

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C12N 15/31, 15/82, 15/10, 1/21, 5/10, A01H 5/00, C07K 14/24, A01N 63/02 (11) International Publication Number:

WO 99/42589

A2

US

(43) International Publication Date:

26 August 1999 (26.08.99)

(21) International Application Number:

PCT/EP99/01015

(22) International Filing Date:

18 February 1999 (18.02.99)

(30) Priority Data:

09/027,080

20 February 1998 (20.02.98)

60/116,439

20 January 1999 (20.01.99) US

(71) Applicant (for all designated States except AT US): NOVAR-TIS AG [CH/CH]; Schwarzwaldallee 215, D-4058 Basel (CH).

(71) Applicant (for AT only): NOVARTIS-ERFINDUNGEN VER-WALTUNGSGESELLSCHAFT MBH [AT/AT]; Brunner Strasse 59, A-1235 Vienna (AT).

(75) Inventors/Applicants (for US only): KRAMER, Vance, Cary [US/US]; 608 Dana Court, Hillsborough, NC 27278 (US). MORGAN, Michael, Kent [US/US]; 5805 Garrett Road, Durham, NC 27707 (US). ANDERSON, Arne, Robert [US/US]; 1005 Green-Pace Road, Zebulon, NC 27597 (US). HART, Hope, Prim [US/US]; 4106 Planters Glen Court, Fuquay-Varina, NC 26526 (US). Warren, Gregory, Wayne [US/US]; 324 Bond Lake Drive, Cary, NC 27513 (US). DUNN, Martha, M. [US/US]; 6201 Oakview Court,

Hillsborough, NC 27278 (US). CHEN, Jeng, Shong [-/US]; 302 Orchard Lane, Chapel Hill, NC 27514 (US).

(74) Agent: BECKER, Konrad; Novartis AG, Corporate Intellectual Property, Patent & Trademark Dept., CH-4002 Basel (CH).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: INSECTICIDAL TOXINS FROM PHOTORHABDUS

(57) Abstract

Novel nucleic acid sequences isolated from Photorhabdus luminescens, whose expression results in novel insecticidal toxins, are disclosed herein. The invention also discloses compositions and formulations containing the insecticidal toxins that are capable of controlling insect pests. The invention is further drawn to methods of making the toxins and to methods of using the nucleotide sequences, for example in microorganisms to control insect pests or in transgenic plants to confer insect resistance.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
ВА	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
вв	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	1E	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
СН	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

WO 99/42589 - 1 -

INSECTICIDAL TOXINS FROM PHOTORHABDUS

The invention relates to novel toxins from *Photorhabdus luminescens*, nucleic acid sequences whose expression results in said toxins, and methods of making and methods of using the toxins and corresponding nucleic acid sequences to control insects.

Insect pests are a major cause of crop losses. Solely in the US, about \$7.7 billion are lost every year due to infestation by various genera of insects. In addition to losses in field crops, insect pests are also a burden to vegetable and fruit growers, to producers of ornamental flowers, and they are a nuisance to gardeners and home owners.

Insect pests are mainly controlled by intensive applications of chemical insecticides, which are active through inhibition of insect growth, prevention of insect feeding or reproduction, or death of the insects. Good insect control can thus be reached, but these chemicals can sometimes also affect other, beneficial insects. Another problem resulting from the wide use of chemical pesticides is the appearance of resistant insect varieties. This has been partially alleviated by various resistance management strategies, but there is an increasing need for alternative pest control agents. Biological insect control agents, such as Bacillus thuringiensis strains expressing insecticidal toxins like d-endotoxins, have also been applied with satisfactory results, offering an alternative or a complement to chemical insecticides. Recently, the genes coding for some of these d-endotoxins have been isolated and their expression in heterologous hosts have been shown to provide another tool for the control of economically important insect pests. In particular, the expression of insecticidal toxins in transgenic plants, such as Bacillus thuringiensis dendotoxins, has provided efficient protection against selected insect pests, and transgenic plants expressing such toxins have been commercialized, allowing farmers to reduce applications of chemical insect control agents. Yet, even in this case, the development of resistance remains a possibility and only a few specific insect pests are controllable. Consequently, there remains a long-felt but unfulfilled need to discover new and effective insect control agents that provide an economic benefit to farmers and that are environmentally acceptable.

The present invention addresses the need for novel insect control agents. Particularly needed are control agents that are targeted to economically important insect pests and that efficiently control insect strains resistant to existing insect control agents.

-2-

Furthermore, agents whose application minimizes the burden on the environment are desirable.

In the search of novel insect control agents, certain classes of nematodes from the genera *Heterorhabdus* and *Steinernema* are of particular interest because of their insecticidal properties. They kill insect larvae and their offspring feed in the dead larvae. Indeed, the insecticidal activity is due to symbiotic bacteria living in the nematodes. These symbiotic bacteria are *Photorhabdus* in the case of *Heterorhabdus* and *Xenorhabdus* in the case of *Steinernema*.

The present invention is drawn to nucleic acid sequences isolated from *Photorhabdus luminescens*, and sequences substantially similar thereto, whose expression results in toxins that are highly toxic to economically important insect pests, particularly insect pests that infest plants. The invention is further drawn to the toxins resulting from the expression of the nucleic acid sequences, and to compositions and formulations containing the toxins, which are capable of inhibiting the ability of insect pests to survive, grow or reproduce, or of limiting insect-related damage or loss in crop plants. The invention is further drawn to a method of making the toxins and to methods of using the nucleic acid sequences, for example in microorganisms to control insects or in transgenic plants to confer insect resistance, and to a method of using the toxins, and compositions and formulations comprising the toxins, for example applying the toxins or compositions or formulations to insect-infested areas, or to prophylactically treat insect-susceptible areas or plants to confer protection or resistance to the insects.

The novel toxins are highly active against insects. For example, a number of economically important insect pests, such as the Lepidopterans *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Manduca sexta* (Tobacco Hornworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm), as well as the Coleopterans *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle) can be controlled by one or more of the toxins. The toxins can be used in multiple insect control strategies, resulting in maximal efficiency with minimal impact on the environment.

According to one aspect, the present invention provides an isolated nucleic acid molecule comprising: (a) a nucleotide sequence substantially similar to a nucleotide

sequence selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 15,171-18,035 of SEQ ID NO:11, and nucleotides 31,393-35,838 of SEQ ID NO:11; (b) a nucleotide sequence comprising nucleotides 23,768-31,336 of SEQ ID NO:11; or (c) a nucleotide sequence isocoding with the nucleotide sequence of (a) or (b); wherein expression of the nucleic acid molecule results in at least one toxin that is active against insects.

In one embodiment of this aspect, the nucleotide sequence is isocoding with a nucleotide sequence substantially similar to nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1. Preferably, the nucleotide sequence is substantially similar to nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1. More preferably, the nucleotide sequence encodes an amino acid sequence selected from the group consisting of SEQ ID NO:2-6. Most preferably, the nucleotide sequence comprises nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1.

In another embodiment of this aspect, the nucleotide sequence is isocoding with a nucleotide sequence substantially similar to nucleotides 15,171-18,035 of SEQ ID NO:11. Preferably, the nucleotide sequence is substantially similar to nucleotides 15,171-18,035 of SEQ ID NO:11. More preferably, the nucleotide sequence encodes the amino acid sequence set forth in SEQ ID NO:12. Most preferably, the nucleotide sequence comprises nucleotides 15,171-18,035 of SEQ ID NO:11.

In still another embodiment of this aspect, the nucleotide sequence is isocoding with a nucleotide sequence substantially similar to nucleotides 31,393-35,838 of SEQ ID NO:11. Preferably, the nucleotide sequence is substantially similar to nucleotides 31,393-35,838 of SEQ ID NO:11. More preferably, the nucleotide sequence encodes the amino acid sequence set forth in SEQ ID NO:14. Most preferably, the nucleotide sequence comprises nucleotides 31,393-35,838 of SEQ ID NO:11.

WO 99/42589 PCT/EP99/01015

- 4 -

In yet another embodiment of this aspect, the nucleotide sequence encodes the amino acid sequence set forth in SEQ ID NO:13, and preferably comprises nucleotides 23,768-31,336 of SEQ ID NO:11.

In one embodiment, the nucleotide sequence of the invention comprises the approximately 9.7 kb DNA fragment harbored in *E. coli* strain DH5a, designated as NRRL accession number B-21835.

In another embodiment, the nucleotide sequence of the invention comprises the approximately 38 kb DNA fragment harbored in *E. coli* strain DH5a, designated as NRRL accession number B-30077.

In still another embodiment, the nucleotide sequence of the invention comprises the approximately 22.2 kb DNA fragment harbored in *E. coli* strain DH5a, designated as NRRL accession number B-30078.

According to one embodiment of the invention, the toxins resulting from expression of the nucleic acid molecules of the invention have activity against Lepidopteran insects. Preferably, according to this embodiment, the toxins have activity against *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm).

According to another embodiment of the invention, the toxins resulting from expression of the nucleic acid molecule of the invention have activity against Lepidopteran and Coleopteran insects. Preferably, according to this embodiment, the toxins have insecticidal activity against *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle).

In another aspect, the present invention provides an isolated nucleic acid molecule comprising a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of a nucleotide sequence selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 15,171-18,035 of SEQ ID NO:11, and nucleotides 31,393-35,838 of SEQ ID NO:11, wherein expression of the nucleic acid molecule results in at least one toxin that is active against insects.

In one embodiment of this aspect, the isolated nucleic acid molecule of the invention comprises a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1.

In another embodiment of this aspect, the isolated nucleic acid molecule of the invention comprises a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of nucleotides 15,171-18,035 of SEQ ID NO:11.

In still another embodiment of this aspect, the isolated nucleic acid molecule of the invention comprises a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of nucleotides 31,393-35,838 of SEQ ID NO:11.

In a further aspect, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence from *Photorhabdus luminescens* selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 66-1898 of SEQ ID NO:11, nucleotides 2416-9909 of SEQ ID NO:11, the complement of nucleotides 2817-3395 of SEQ ID NO:11, nucleotides 9966-14,633 of SEQ ID NO:11, nucleotides 14,699-15,007 of SEQ ID NO:11, nucleotides 15,171-18,035 of SEQ ID NO:11, the complement of nucleotides 17,072-17,398 of SEQ ID NO:11, the complement of nucleotides 18,235-19,167 of SEQ ID NO:11, the complement of nucleotides 20,217-20,963 of SEQ ID NO:11, the complement of nucleotides 22,172-23,086 of SEQ ID NO:11, nucleotides 23,768-31,336 of SEQ ID NO:11, nucleotides 31,393-35,838 of SEQ ID NO:11, the complement of nucleotides 35,383-35,709 of SEQ ID NO:11, the complement of nucleotides 36,032-36,661 of SEQ ID NO:11, and the complement of nucleotides 36,654-37,781 of SEQ ID NO:11.

The present invention also provides a chimeric gene comprising a heterologous promoter sequence operatively linked to the nucleic acid molecule of the invention. Further, the present invention provides a recombinant vector comprising such a chimeric gene. Still further, the present invention provides a host cell comprising such a chimeric gene. A host cell according to this aspect of the invention may be a bacterial cell, a yeast cell, or a plant

cell, preferably a plant cell. Even further, the present invention provides a plant comprising such a plant cell. Preferably, the plant is maize.

In yet another aspect, the present invention provides toxins produced by the expression of DNA molecules of the present invention.

According to one embodiment, the toxins of the invention have activity against Lepidopteran insects, preferably against *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm).

According to another embodiment, the toxins of the invention have activity against Lepidopteran and Coleopteran insects, preferably against *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle).

In one embodiment, the toxins are produced by the *E. coli* strain designated as NRRL accession number B-21835.

In another embodiment, the toxins are produced by *E. coli* strain designated as NRRL accession number B-30077.

In still another embodiment, the toxins are produced by *E. coli* strain designated as NRRL accession number B-30078.

In one embodiment, a toxin of the invention comprises an amino acid sequence selected from the group consisting of: SEQ ID NOs:2-6.

In another embodiment, a toxin of the invention comprises an amino acid sequence selected from the group consisting of: SEQ ID NOs:12-14.

The present invention also provides a composition comprising an insecticidally effective amount of a toxin according to the invention.

In another aspect, the present invention provides a method of producing a toxin that is active against insects, comprising: (a) obtaining a host cell comprising a chimeric gene, which itself comprises a heterologous promoter sequence operatively linked to the nucleic acid molecule of the invention; and (b) expressing the nucleic acid molecule in the cell, which results in at least one toxin that is active against insects.

In a further aspect, the present invention provides a method of producing an insect-resistant plant, comprising introducing a nucleic acid molecule of the invention into the plant, wherein the nucleic acid molecule is expressible in the plant in an effective amount to control insects. According to one embodiment, the insects are Lepidopteran insects, preferably selected from the group consisting of: *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm). According to another embodiment, the insects are Lepidopteran and Coleopteran insects, preferably selected from the group consisting of: *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa. decimlineata* (Colorado Potato Beetle).

In still a further aspect, the present invention provides a method of controlling insects comprising delivering to the insects an effective amount of a toxin according to the present invention. According to one embodiment, the insects are Lepidopteran insects, preferably selected from the group consisting of: *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm). According to another embodiment, the insects are Lepidopteran and Coleopteran insects, preferably selected from the group consisting of: *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle). Preferably, the toxin is delivered to the insects orally.

Yet another aspect of the present invention is the provision of a method for mutagenizing a nucleic acid molecule according to the present invention, wherein the nucleic acid molecule has been cleaved into population of double-stranded random fragments of a desired size, comprising: (a) adding to the population of double-stranded random fragments one or more single- or double-stranded oligonucleotides, wherein the oligonucleotides each comprise an area of identity and an area of heterology to a double-stranded template polynucleotide; (b) denaturing the resultant mixture of double-stranded

random fragments and oligonucleotides into single-stranded fragments; (c) incubating the resultant population of single-stranded fragments with a polymerase under conditions which result in the annealing of the single-stranded fragments at the areas of identity to form pairs of annealed fragments, the areas of identity being sufficient for one member of a pair to prime replication of the other, thereby forming a mutagenized double-stranded polynucleotide; and (d) repeating the second and third steps for at least two further cycles, wherein the resultant mixture in the second step of a further cycle includes the mutagenized double-stranded polynucleotide from the third step of the previous cycle, and wherein the further cycle forms a further mutagenized double-stranded polynucleotide.

Other aspects and advantages of the present invention will become apparent to those skilled in the art from a study of the following description of the invention and non-limiting examples.

DEFINITIONS

"Activity" of the toxins of the invention is meant that the toxins function as orally active insect control agents, have a toxic effect, or are able to disrupt or deter insect feeding, which may or may not cause death of the insect. When a toxin of the invention is delivered to the insect, the result is typically death of the insect, or the insect does not feed upon the source that makes the toxin available to the insect.

"Associated with / operatively linked" refer to two nucleic acid sequences that are related physically or functionally. For example, a promoter or regulatory DNA sequence is said to be "associated with" a DNA sequence that codes for an RNA or a protein if the two sequences are operatively linked, or situated such that the regulator DNA sequence will affect the expression level of the coding or structural DNA sequence.

A "chimeric gene" is a recombinant nucleic acid sequence in which a promoter or regulatory nucleic acid sequence is operatively linked to, or associated with, a nucleic acid sequence that codes for an mRNA or which is expressed as a protein, such that the regulator nucleic acid sequence is able to regulate transcription or expression of the associated nucleic acid sequence. The regulator nucleic acid sequence of the chimeric gene is not normally operatively linked to the associated nucleic acid sequence as found in nature.

A "coding sequence" is a nucleic acid sequence that is transcribed into RNA such as mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. Preferably the RNA is then translated in an organism to produce a protein.

To "control" insects means to inhibit, through a toxic effect, the ability of insect pests to survive, grow, feed, and/or reproduce, or to limit insect-related damage or loss in crop plants. To "control" insects may or may not mean killing the insects, although it preferably means killing the insects.

To "deliver" a toxin means that the toxin comes in contact with an insect, resulting in toxic effect and control of the insect. The toxin can be delivered in many recognized ways, e.g., orally by ingestion by the insect or by contact with the insect via transgenic plant expression, formulated protein composition(s), sprayable protein composition(s), a bait matrix, or any other art-recognized toxin delivery system.

"Expression cassette" as used herein means a nucleic acid sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operably linked to the nucleotide sequence of interest which is operably linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The expression cassette comprising the nucleotide sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette may also be one which is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. Typically, however, the expression cassette is heterologous with respect to the host, i.e., the particular nucleic acid sequence of the expression cassette does not occur naturally in the host cell and must have been introduced into the host cell or an ancestor of the host cell by a transformation event. The expression of the nucleotide sequence in the expression cassette may be under the control of a constitutive promoter or of an inducible promoter which initiates transcription only when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism, such as a plant, the promoter can also be specific to a particular tissue, or organ, or stage of development.

A "gene" is a defined region that is located within a genome and that, besides the aforementioned coding nucleic acid sequence, comprises other, primarily regulatory, nucleic acid sequences responsible for the control of the expression, that is to say the transcription and translation, of the coding portion. A gene may also comprise other 5' and 3'

untranslated sequences and termination sequences. Further elements that may be present are, for example, introns.

"Gene of interest" refers to any gene which, when transferred to a plant, confers upon the plant a desired characteristic such as antibiotic resistance, virus resistance, insect resistance, disease resistance, or resistance to other pests, herbicide tolerance, improved nutritional value, improved performance in an industrial process or altered reproductive capability. The "gene of interest" may also be one that is transferred to plants for the production of commercially valuable enzymes or metabolites in the plant.

A "heterologous" nucleic acid sequence is a nucleic acid sequence not naturally associated with a host cell into which it is introduced, including non-naturally occurring multiple copies of a naturally occurring nucleic acid sequence.

A "homologous" nucleic acid sequence is a nucleic acid sequence naturally associated with a host cell into which it is introduced.

"Homologous recombination" is the reciprocal exchange of nucleic acid fragments between homologous nucleic acid molecules.

"Insecticidal" is defined as a toxic biological activity capable of controlling insects, preferably by killing them.

A nucleic acid sequence is "isocoding with" a reference nucleic acid sequence when the nucleic acid sequence encodes a polypeptide having the same amino acid sequence as the polypeptide encoded by the reference nucleic acid sequence.

An "isolated" nucleic acid molecule or an isolated enzyme is a nucleic acid molecule or enzyme that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated nucleic acid molecule or enzyme may exist in a purified form or may exist in a non-native environment such as, for example, a recombinant host cell.

A "nucleic acid molecule" or "nucleic acid sequence" is a linear segment of single- or double-stranded DNA or RNA that can be isolated from any source. In the context of the present invention, the nucleic acid molecule is preferably a segment of DNA.

"ORF" means open reading frame.

A "plant" is any plant at any stage of development, particularly a seed plant.

A "plant cell" is a structural and physiological unit of a plant, comprising a protoplast and a cell wall. The plant cell may be in form of an isolated single cell or a cultured cell, or as a part of higher organized unit such as, for example, plant tissue, a plant organ, or a whole plant.

"Plant cell culture" means cultures of plant units such as, for example, protoplasts, cell culture cells, cells in plant tissues, pollen, pollen tubes, ovules, embryo sacs, zygotes and embryos at various stages of development.

"Plant material" refers to leaves, stems, roots, flowers or flower parts, fruits, pollen, egg cells, zygotes, seeds, cuttings, cell or tissue cultures, or any other part or product of a plant.

A "plant organ" is a distinct and visibly structured and differentiated part of a plant such as a root, stem, leaf, flower bud, or embryo.

"Plant tissue" as used herein means a group of plant cells organized into a structural and functional unit. Any tissue of a plant *in planta* or in culture is included. This term includes, but is not limited to, whole plants, plant organs, plant seeds, tissue culture and any groups of plant cells organized into structural and/or functional units. The use of this term in conjunction with, or in the absence of, any specific type of plant tissue as listed above or otherwise embraced by this definition is not intended to be exclusive of any other type of plant tissue.

A "promoter" is an untranslated DNA sequence upstream of the coding region that contains the binding site for RNA polymerase II and initiates transcription of the DNA. The promoter region may also include other elements that act as regulators of gene expression.

A "protoplast" is an isolated plant cell without a cell wall or with only parts of the cell wall.

"Regulatory elements" refer to sequences involved in controlling the expression of a nucleotide sequence. Regulatory elements comprise a promoter operably linked to the nucleotide sequence of interest and termination signals. They also typically encompass sequences required for proper translation of the nucleotide sequence.

In its broadest sense, the term "substantially similar", when used herein with respect to a nucleotide sequence, means a nucleotide sequence corresponding to a reference nucleotide sequence, wherein the corresponding sequence encodes a polypeptide having substantially the same structure and function as the polypeptide encoded by the reference nucleotide sequence, e.g. where only changes in amino acids not affecting the polypeptide function occur. Desirably the substantially similar nucleotide sequence encodes the polypeptide encoded by the reference nucleotide sequence. The percentage of identity between the substantially similar nucleotide sequence and the reference nucleotide sequence desirably is at least 80%, more desirably at least 85%, preferably at least 90%, more preferably at least 95%, still more preferably at least 99%. A nucleotide sequence

"substantially similar" to reference nucleotide sequence hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

"Synthetic" refers to a nucleotide sequence comprising structural characters that are not present in the natural sequence. For example, an artificial sequence that resembles more closely the G+C content and the normal codon distribution of dicot and/or monocot genes is said to be synthetic.

"Transformation" is a process for introducing heterologous nucleic acid into a host cell or organism. In particular, "transformation" means the stable integration of a DNA molecule into the genome of an organism of interest.

"Transformed / transgenic / recombinant" refer to a host organism such as a bacterium or a plant into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be stably integrated into the genome of the host or the nucleic acid molecule can also be present as an extrachromosomal molecule. Such an extrachromosomal molecule can be auto-replicating. Transformed cells, tissues, or plants are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof. A "non-transformed", "non-transgenic", or "non-recombinant" host refers to a wild-type organism, e.g., a bacterium or plant, which does not contain the heterologous nucleic acid molecule.

Nucleotides are indicated by their bases by the following standard abbreviations: adenine (A), cytosine (C), thymine (T), and guanine (G). Amino acids are likewise indicated by the following standard abbreviations: alanine (Ala; A), arginine (Arg; R), asparagine (Asn; N), aspartic acid (Asp; D), cysteine (Cys; C), glutamine (Gln; Q), glutamic acid (Glu; E), glycine (Gly; G), histidine (His; H), isoleucine (Ile; I), leucine (Leu; L), lysine (Lys; K), methionine (Met; M), phenylalanine (Phe; F), proline (Pro; P), serine (Ser; S), threonine (Thr; T), tryptophan (Trp; W), tyrosine (Tyr; Y), and valine (Val; V). Furthermore, (Xaa; X) represents any amino acid.

WO 99/42589 PCT/EP99/01015

BRIEF DESCRIPTION OF THE SEQUENCES IN THE SEQUENCE LISTING

SEQ ID NO:1 is the sequence of the approximately 9.7 kb DNA fragment comprised in pCIB9359-7 which comprises the following ORFs at the specified nucleotide positions:

<u>Name</u>	Start	<u>End</u>
orf1	412	1665
orf2	1686	2447
orf3	2758	3318
orf4	3342	4118
orf5	4515	9269

SEQ ID NO:2 is the sequence of the ~46.4 kDa protein encoded by orf1 of SEQ ID NO:1.

SEQ ID NO:3 is the sequence of the ~28.1 kDa protein encoded by orf2 of SEQ ID NO:1.

SEQ ID NO:4 is the sequence of the ~20.7 kDa protein encoded by orf3 of SEQ ID NO:1.

SEQ ID NO:5 is the sequence of the ~28.7 kDa protein encoded by orf4 of SEQ ID NO:1.

SEQ ID NO:6 is the sequence of the ~176 kDa protein encoded by orf5 of SEQ ID NO:1.

SEQ ID NOs:7-10 are oligonucleotides.

SEQ ID NO:11 is the sequence of the approximately 38 kb DNA fragment comprised in pNOV2400, which comprises the following ORFs at the specified nucleotide positions (descending numbers and "C" indicates that the ORF is on the complementary strand):

<u>Name</u>	Start	<u>End</u>	
orf7	66	1898	(partial sequence)
hph3	2416	9909	
orf18	3395	2817	С
orf4	9966	14,633	
orf19	14,699	15,007	
orf5	15,1 71	18,035	
orf22	17,398	17,072	С
orf10	19,167	18,235	С
orf14	20,116	19,385	С
orf13	20,963	20,217	С
orf11	23,086	22,172	С
hph2	23,768	31,336	
orf2	31,393	35,838	

WO 99/42589 PCT/EP99/01015

- 14 -

orf21 35,709 35,383 C orf16 36,661 36,032 C orf8 37,781 36,654 C

SEQ ID NO:11 also includes the following restriction sites, some of which are used in the subcloning steps set forth in Example 17:

Restriction Site	Nucleotide Position(s)
Accll	2835
<i>Bam</i> HI	18,915
<i>Bsm</i> Bl	11,350
Bst11071	29,684
<i>Eag</i> l	13,590; 31,481
Eco721	34,474
Mlui	2444; 5116; 9327; 26,204
Notl	13,589
<i>Pac</i> l	9915; 23,353; 37,888
Pvul	8816
Sapl	35,248
SexAI	28,946
Sgfl	8815
Spel	2157; 3769; 7831; 11,168
Sphi	755
Stul	35,690
Tth1111	21,443

SEQ ID NO:12 is the sequence of the protein encoded by orf5 of SEQ ID NO:11.

SEQ ID NO:13 is the sequence of the protein encoded by hph2 of SEQ ID NO:11.

SEQ ID NO:14 is the sequence of the protein encoded by orf2 of SEQ ID NO:11.

SEQ ID NOs:15-22 are oligonucleotides.

DEPOSITS

The following material has been deposited with the Agricultural Research Service, Patent Culture Collection (NRRL), 1815 North University Street, Peoria, Illinois 61604, under the terms of the Budapest Treaty on the International Recognition of the Deposit of

Microorganisms for the Purposes of Patent Procedure. All restrictions on the availability of the deposited material will be irrevocably removed upon the granting of a patent.

Clone	Accession Number	Date of Deposit
pCIB9359-7	NRRL B-21835	September 17, 1997
pNOV2400	NRRL B-30077	December 3, 1998
pNOV1001	NRRL B-30078	December 3, 1998

Novel Nucleic Acid Sequences whose Expression Results in Insecticidal Toxins

This invention relates to nucleic acid sequences whose expression results in novel toxins, and to the making and using of the toxins to control insect pests. The nucleic acid sequences are derived from Photorhabdus luminescens, a member of Enterobacteriaceae family. P. luminescens is a symbiotic bacterium of nematodes of the genus Heterorhabditis. The nematodes colonize insect larva, kill them, and their offspring feed on the dead larvae. The insecticidal activity is actually produced by the symbiotic P. luminescens bacteria. The inventors are the first to isolate the nucleic acid sequences of the present invention from P. luminescens (ATCC strain number 29999). The expression of the nucleic acid sequences of the present invention results in toxins that can be used to control Lepidopteran insects such as Plutella xylostella (Diamondback Moth), Trichoplusia ni (Cabbage Looper), Ostrinia nubilalis (European Corn Borer), Heliothis virescens (Tobacco Budworm), Helicoverpa zea (Corn Earworm), Manduca sexta (Tobacco Hornworm), Spodoptera exigua (Beet Armyworm), and Spodoptera frugiperda (Fall Armyworm), as well as Coleopteran insects such as Diabrotica virgifera virgifera (Western Corn Rootworm), Diabrotica undecimpunctata howardi (Southern Corn Rootworm), Diabrotica longicornis barberi (Northern Corn Rootworm), and Leptinotarsa decimlineata (Colorado Potato Beetle).

In one preferred embodiment, the invention encompasses an isolated nucleic acid molecule comprising a nucleotide sequence substantially similar to the approximately 9.7 kb nucleic acid sequence set forth in SEQ ID NO:1, whose expression results in insect control activity (further illustrated in Examples 1-11). Five open reading frames (ORFs) are present in the nucleic acid sequence set forth in SEQ ID NO:1, coding for proteins of predicted sizes 45 kDa, 28 kDa, 21 kDA, 29 kDa, and 176 kDa. The five ORFs are arranged in an operon-like structure. When expressed in a heterologous host, the ~ 9.7 kb DNA fragment from *P*.

WO 99/42589 PCT/EP99/01015

- 16 -

luminescens results in insect control activity against Lepidopterans such as *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm), showing that expression of the ~ 9.7 kb nucleotide sequence set forth in SEQ ID NO:1 is necessary and sufficient for such insect control activity. In a preferred embodiment, the invention encompasses a DNA molecule, whose expression results in an insecticidal toxin, which is deposited in the *E. coli* strain pCIB9359-7 (NRRL accession number B-21835).

In another preferred embodiment, the invention encompasses an isolated nucleic acid molecule comprising a nucleotide sequence substantially similar to the approximately 38 kb nucleic acid fragment set forth in SEQ ID NO:11 and deposited in the E. coli strain pNOV2400 (NRRL accession number B-30077), whose expression results in insect control activity (see Examples 12-18). In a more preferred embodiment, the invention encompasses an isolated nucleic acid molecule comprising a nucleotide sequence substantially similar to the ~ 22 kb DNA fragment deposited in the E. coli strain pNOV1001 (NRRL accession number B-30078), whose expression results in insect control activity. In a most preferred embodiment, the invention encompasses isolated nucleic acid molecules comprising nucleotide sequences substantially similar to the three ORFs corresponding to nucleotides 23,768-31,336 (hph2), 31,393-35,838 (orf2), and 15,171-18,035 (orf5) of the DNA fragment set forth in SEQ ID NO:11, as well as the proteins encoded thereby. When co-expressed in a heterologous host, these three ORFs result in insect control activity against Lepidopterans such as Plutella xylostella (Diamondback Moth), Ostrinia nubilalis (European Corn Borer), and Manduca sexta (Tobacco Hornworm), as well as against Coleopterans such as Diabrotica virgifera virgifera (Western Corn Rootworm), Diabrotica undecimpunctata howardi (Southern Corn Rootworm), and Leptinotarsa decimlineata (Colorado Potato Beetle), showing that co-expression of these three ORFs (hph2, orf2, and orf5) is necessary and sufficient for such insect control activity.

The present invention also encompasses recombinant vectors comprising the nucleic acid sequences of this invention. In such vectors, the nucleic acid sequences are preferably comprised in expression cassettes comprising regulatory elements for expression of the nucleotide sequences in a host cell capable of expressing the nucleotide sequences. Such regulatory elements usually comprise promoter and termination signals and preferably also

comprise elements allowing efficient translation of polypeptides encoded by the nucleic acid sequences of the present invention. Vectors comprising the nucleic acid sequences are usually capable of replication in particular host cells, preferably as extrachromosomal molecules, and are therefore used to amplify the nucleic acid sequences of this invention in the host cells. In one embodiment, host cells for such vectors are microorganisms, such as bacteria, in particular E.coli. In another embodiment, host cells for such recombinant vectors are endophytes or epiphytes. A preferred host cell for such vectors is a eukaryotic cell, such as a yeast, a plant cell, or an insect cell. Plant cells such as maize cells are most preferred host cells. In another preferred embodiment, such vectors are viral vectors and are used for replication of the nucleotide sequences in particular host cells, e.g. insect cells or plant cells. Recombinant vectors are also used for transformation of the nucleotide sequences of this invention into host cells, whereby the nucleotide sequences are stably integrated into the DNA of such host cells. In one, such host cells are prokaryotic cells. In a preferred embodiment, such host cells are eukaryotic cells, such as yeast cells, insect cells, or plant cells. In a most preferred embodiment, the host cells are plant cells, such as maize cells.

In preferred embodiments, the insecticidal toxins of the invention each comprise at least one polypeptide encoded by a nucleotide sequence of the invention. In another preferred embodiment, the insecticidal toxins are produced from a purified strain of *P. luminescens*, such the strain with ATTC accession number 29999. The toxins of the present invention have insect control activity when tested against insect pests in bioassays; and these properties of the insecticidal toxins are further illustrated in Examples 1-18. The insecticidal toxins desribed in the present invention are further characterized in that their molecular weights are larger than 6,000, as found by size fractionation experiments. The insecticidal toxins retain full insectidical activity after being stored at 4°C for 2 weeks. One is also shown to retain its full insecticidal activity after being freeze-dried and stored at 22°C for 2 weeks. However, the insecticidal toxins of the invention lose their insecticidal activity after incubation for 5 minutes at 100°C.

In further embodiments, the nucleotide sequences of the invention can be modified by incorporation of random mutations in a technique known as *in-vitro* recombination or DNA shuffling. This technique is described in Stemmer et al., Nature 370: 389-391 (1994) and US Patent 5,605,793, which are incorporated herein by reference. Millions of mutant copies of a nucleotide sequence are produced based on an original nucleotide sequence of

this invention and variants with improved properties, such as increased insecticidal activity, enhanced stability, or different specificity or range of target insect pests are recovered. The method encompasses forming a mutagenized double-stranded polynucleotide from a template double-stranded polynucleotide comprising a nucleotide sequence of this invention, wherein the template double-stranded polynucleotide has been cleaved into double-stranded-random fragments of a desired size, and comprises the steps of adding to the resultant population of double-stranded random fragments one or more single or double-stranded oligonucleotides, wherein said oligonucleotides comprise an area of identity and an area of heterology to the double-stranded template polynucleotide; denaturing the resultant mixture of double-stranded random fragments and oligonucleotides into single-stranded fragments; incubating the resultant population of single-stranded fragments with a polymerase under conditions which result in the annealing of said singlestranded fragments at said areas of identity to form pairs of annealed fragments, said areas of identity being sufficient for one member of a pair to prime replication of the other, thereby forming a mutagenized double-stranded polynucleotide; and repeating the second and third steps for at least two further cycles, wherein the resultant mixture in the second step of a further cycle includes the mutagenized double-stranded polynucleotide from the third step of the previous cycle, and the further cycle forms a further mutagenized double-stranded polynucleotide. In a preferred embodiment, the concentration of a single species of doublestranded random fragment in the population of double-stranded random fragments is less than 1% by weight of the total DNA. In a further preferred embodiment, the template double-stranded polynucleotide comprises at least about 100 species of polynucleotides. In another preferred embodiment, the size of the double-stranded random fragments is from about 5 bp to 5 kb. In a further preferred embodiment, the fourth step of the method comprises repeating the second and the third steps for at least 10 cycles.

Expression of the Nucleotide Sequences in Heterologous Microbial Hosts

As biological insect control agents, the insecticidal toxins are produced by expression of the nucleotide sequences in heterologous host cells capable of expressing the nucleotide sequences. In a first embodiment, *P. luminescens* cells comprising modifications of at least one nucleotide sequence of this invention at its chromosomal location are described. Such modifications encompass mutations or deletions of existing regulatory elements, thus leading to altered expression of the nucleotide sequence, or the incorporation of new regulatory elements controlling the expression of the nucleotide sequence. In another

embodiment, additional copies of one or more of the nucleotide sequences are added to *P. luminescens* cells either by insertion into the chromosome or by introduction of extrachromosomally replicating molecules containing the nucleotide sequences.

In another embodiment, at least one of the nucleotide sequences of the invention is inserted into an appropriate expression cassette, comprising a promoter and termination signals. Expression of the nucleotide sequence is constitutive, or an inducible promoter responding to various types of stimuli to initiate transcription is used. In a preferred embodiment, the cell in which the toxin is expressed is a microorganism, such as a virus, a bacteria, or a fungus. In a preferred embodiment, a virus, such as a baculovirus, contains a nucleotide sequence of the invention in its genome and expresses large amounts of the corresponding insecticidal toxin after infection of appropriate eukaryotic cells that are suitable for virus replication and expression of the nucleotide sequence. The insecticidal toxin thus produced is used as an insecticidal agent. Alternatively, baculoviruses engineered to include the nucleotide sequence are used to infect insects *in-vivo* and kill them either by expression of the insecticidal toxin or by a combination of viral infection and expression of the insecticidal toxin.

Bacterial cells are also hosts for the expression of the nucleotide sequences of the invention. In a preferred embodiment, non-pathogenic symbiotic bacteria, which are able to live and replicate within plant tissues, so-called endophytes, or non-pathogenic symbiotic bacteria, which are capable of colonizing the phyllosphere or the rhizosphere, so-called epiphytes, are used. Such bacteria include bacteria of the genera Agrobacterium, Alcaligenes, Azospirillum, Azotobacter, Bacillus, Clavibacter, Enterobacter, Erwinia, Flavobacter, Klebsiella, Pseudomonas, Rhizobium, Serratia, Streptomyces and Xanthomonas. Symbiotic fungi, such as Trichoderma and Gliocladium are also possible hosts for expression of the inventive nucleotide sequences for the same purpose.

Techniques for these genetic manipulations are specific for the different available hosts and are known in the art. For example, the expression vectors pKK223-3 and pKK223-2 can be used to express heterologous genes in *E. coli*, either in transcriptional or translational fusion, behind the *tac or trc* promoter. For the expression of operons encoding multiple ORFs, the simplest procedure is to insert the operon into a vector such as pKK223-3 in transcriptional fusion, allowing the cognate ribosome binding site of the heterologous genes to be used. Techniques for overexpression in gram-positive species such as *Bacillus* are also known in the art and can be used in the context of this invention (Quax *et al. In.:*

- 20 -

Industrial Microorganisms: Basic and Applied Molecular Genetics, *Eds.* Baltz *et al.*, American Society for Microbiology, Washington (1993)). Alternate systems for overexpression rely for example, on yeast vectors and include the use of *Pichia*, *Saccharomyces* and *Kluyveromyces* (Sreekrishna, *In*: Industrial microorganisms: basic and applied molecular genetics, Baltz, Hegeman, and Skatrud *eds.*, American Society for Microbiology, Washington (1993); Dequin & Barre, Biotechnology 12:173-177 (1994); van den Berg *et al.*, Biotechnology 8:135-139 (1990)).

In another preferred embodiment, at least one of the described nucleotide sequences is transferred to and expressed in *Pseudomonas fluorescens* strain CGA267356 (described in the published application EU 0 472 494 and in WO 94/01561) which has biocontrol characteristics. In another preferred embodiment, a nucleotide sequence of the invention is transferred to *Pseudomonas aureofaciens* strain 30-84 which also has biocontrol characteristics. Expression in heterologous biocontrol strains requires the selection of vectors appropriate for replication in the chosen host and a suitable choice of promoter. Techniques are well known in the art for expression in gram-negative and grampositive bacteria and fungi.

Expression of the Nucleotide Sequences in Plant Tissue

In a particularly preferred embodiment, at least one of the insecticidal toxins of the invention is expressed in a higher organism, e.g., a plant. In this case, transgenic plants expressing effective amounts of the toxins protect themselves from insect pests. When the insect starts feeding on such a transgenic plant, it also ingests the expressed toxins. This will deter the insect from further biting into the plant tissue or may even harm or kill the insect. A nucleotide sequence of the present invention is inserted into an expression cassette, which is then preferably stably integrated in the genome of said plant. In another preferred embodiment, the nucleotide sequence is included in a non-pathogenic self-replicating virus. Plants transformed in accordance with the present invention may be monocots or dicots and include, but are not limited to, maize, wheat, barley, rye, sweet potato, bean, pea, chicory, lettuce, cabbage, cauliflower, broccoli, turnip, radish, spinach, asparagus, onion, garlic, pepper, celery, squash, pumpkin, hemp, zucchini, apple, pear, quince, melon, plum, cherry, peach, nectarine, apricot, strawberry, grape, raspberry, blackberry, pineapple, avocado, papaya, mango, banana, soybean, tomato, sorghum, sugarcane, sugarbeet, sunflower, rapeseed, clover, tobacco, carrot, cotton, alfalfa, rice,

- 21 -

potato, eggplant, cucumber, *Arabidopsis*, and woody plants such as coniferous and deciduous trees.

Once a desired nucleotide sequence has been transformed into a particular plant species, it may be propagated in that species or moved into other varieties of the same species, particularly including commercial varieties, using traditional breeding techniques.

A nucleotide sequence of this invention is preferably expressed in transgenic plants, thus causing the biosynthesis of the corresponding toxin in the transgenic plants. In this way, transgenic plants with enhanced resistance to insects are generated. For their expression in transgenic plants, the nucleotide sequences of the invention may require modification and optimization. Although in many cases genes from microbial organisms can be expressed in plants at high levels without modification, low expression in transgenic plants may result from microbial nucleotide sequences having codons that are not preferred in plants. It is known in the art that all organisms have specific preferences for codon usage, and the codons of the nucleotide sequences described in this invention can be changed to conform with plant preferences, while maintaining the amino acids encoded thereby. Furthermore, high expression in plants is best achieved from coding sequences that have at least 35% about GC content, preferably more than about 45%, more preferably more than about 50%, and most preferably more than about 60%. Microbial nucleotide sequences which have low GC contents may express poorly in plants due to the existence of ATTTA motifs which may destabilize messages, and AATAAA motifs which may cause inappropriate polyadenylation. Although preferred gene sequences may be adequately expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray et al. Nucl. Acids Res. 17: 477-498 (1989)). In addition, the nucleotide sequences are screened for the existence of illegitimate splice sites that may cause message truncation. All changes required to be made within the nucleotide sequences such as those described above are made using well known techniques of site directed mutagenesis, PCR, and synthetic gene construction using the methods described in the published patent applications EP 0 385 962 (to Monsanto), EP 0 359 472 (to Lubrizol, and WO 93/07278 (to Ciba-Geigy).

For efficient initiation of translation, sequences adjacent to the initiating methionine may require modification. For example, they can be modified by the inclusion of sequences known to be effective in plants. Joshi has suggested an appropriate consensus for plants

WO 99/42589 PCT/EP99/01015

- 22 -

(NAR <u>15</u>: 6643-6653 (1987)) and Clontech suggests a further consensus translation initiator (1993/1994 catalog, page 210). These consensuses are suitable for use with the nucleotide sequences of this invention. The sequences are incorporated into constructions comprising the nucleotide sequences, up to and including the ATG (whilst leaving the second amino acid unmodified), or alternatively up to and including the GTC subsequent to the ATG (with the possibility of modifying the second amino acid of the transgene).

Expression of the nucleotide sequences in transgenic plants is driven by promoters shown to be functional in plants. The choice of promoter will vary depending on the temporal and spatial requirements for expression, and also depending on the target species. Thus, expression of the nucleotide sequences of this invention in leaves, in ears, in inflorescences (e.g. spikes, panicles, cobs, etc.), in roots, and/or seedlings is preferred. In many cases, however, protection against more than one type of insect pest is sought, and thus expression in multiple tissues is desirable. Although many promoters from dicotyledons have been shown to be operational in monocotyledons and vice versa, ideally are selected for expression in dicotyledonous promoters dicotyledons, and monocotyledonous promoters for expression in monocotyledons. However, there is no restriction to the provenance of selected promoters; it is sufficient that they are operational in driving the expression of the nucleotide sequences in the desired cell.

Preferred promoters that are expressed constitutively include promoters from genes encoding actin or ubiquitin and the CaMV 35S and 19S promoters. The nucleotide sequences of this invention can also be expressed under the regulation of promoters that are chemically regulated. This enables the insecticidal toxins to be synthesized only when the crop plants are treated with the inducing chemicals. Preferred technology for chemical induction of gene expression is detailed in the published application EP 0 332 104 (to Ciba-Geigy) and US patent 5,614,395. A preferred promoter for chemical induction is the tobacco PR-1a promoter.

A preferred category of promoters is that which is wound inducible. Numerous promoters have been described which are expressed at wound sites and also at the sites of phytopathogen infection. Ideally, such a promoter should only be active locally at the sites of infection, and in this way the insecticidal toxins only accumulate in cells which need to synthesize the insecticidal toxins to kill the invading insect pest. Preferred promoters of this kind include those described by Stanford *et al.* Mol. Gen. Genet. <u>215</u>: 200-208 (1989), Xu *et al.* Plant Molec. Biol. <u>22</u>: 573-588 (1993), Logemann *et al.* Plant Cell <u>1</u>: 151-158 (1989),

WO 99/42589 PCT/EP99/01015

- 23 -

Rohrmeier & Lehle, Plant Molec. Biol. <u>22</u>: 783-792 (1993), Firek *et al.* Plant Molec. Biol. <u>22</u>: 129-142 (1993), and Warner *et al.* Plant J. <u>3</u>: 191-201 (1993).

Preferred tissue specific expression patterns include green tissue specific, root specific, stem specific, and flower specific. Promoters suitable for expression in green tissue include many which regulate genes involved in photosynthesis and many of these have been cloned from both monocotyledons and dicotyledons. A preferred promoter is the maize PEPC promoter from the phosphoenol carboxylase gene (Hudspeth & Grula, Plant Molec. Biol. 12: 579-589 (1989)). A preferred promoter for root specific expression is that described by de Framond (FEBS 290: 103-106 (1991); EP 0 452 269 to Ciba-Geigy). A preferred stem specific promoter is that described in US patent 5,625,136 (to Ciba-Geigy) and which drives expression of the maize *trpA* gene.

Especially preferred embodiments of the invention are transgenic plants expressing at least one of the nucleotide sequences of the invention in a root-preferred or root-specific fashion. Further preferred embodiments are transgenic plants expressing the nucleotide sequences in a wound-inducible or pathogen infection-inducible manner.

In addition to the selection of a suitable promoter, constructions for expression of an insecticidal toxin in plants require an appropriate transcription terminator to be attached downstream of the heterologous nucleotide sequence. Several such terminators are available and known in the art (e.g. tm1 from CaMV, E9 from rbcS). Any available terminator known to function in plants can be used in the context of this invention.

Numerous other sequences can be incorporated into expression cassettes described in this invention. These include sequences which have been shown to enhance expression such as intron sequences (e.g. from Adh1 and bronze1) and viral leader sequences (e.g. from TMV, MCMV and AMV).

It may be preferable to target expression of the nucleotide sequences of the present invention to different cellular localizations in the plant. In some cases, localization in the cytosol may be desirable, whereas in other cases, localization in some subcellular organelle may be preferred. Subcellular localization of transgene encoded enzymes is undertaken using techniques well known in the art. Typically, the DNA encoding the target peptide from a known organelle-targeted gene product is manipulated and fused upstream of the nucleotide sequence. Many such target sequences are known for the chloroplast and their functioning in heterologous constructions has been shown. The expression of the

nucleotide sequences of the present invention is also targeted to the endoplasmic reticulum or to the vacuoles of the host cells. Techniques to achieve this are well-known in the art.

Vectors suitable for plant transformation are described elsewhere in this specification. For *Agrobacterium*-mediated transformation, binary vectors or vectors carrying at least one T-DNA border sequence are suitable, whereas for direct gene transfer any vector is suitable and linear DNA containing only the construction of interest may be preferred. In the case of direct gene transfer, transformation with a single DNA species or co-transformation can be used (Schocher *et al.* Biotechnology 4: 1093-1096 (1986)). For both direct gene transfer and *Agrobacterium*-mediated transfer, transformation is usually (but not necessarily) undertaken with a selectable marker which may provide resistance to an antibiotic (kanamycin, hygromycin or methotrexate) or a herbicide (basta). The choice of selectable marker is not, however, critical to the invention.

In another preferred embodiment, a nucleotide sequence of the present invention is directly transformed into the plastid genome. A major advantage of plastid transformation is that plastids are generally capable of expressing bacterial genes without substantial modification, and plastids are capable of expressing multiple open reading frames under control of a single promoter. Plastid transformation technology is extensively described in U.S. Patent Nos. 5,451,513, 5,545,817, and 5,545,818, in PCT application no. WO 95/16783, and in McBride et al. (1994) Proc. Natl. Acad. Sci. USA 91, 7301-7305. The basic technique for chloroplast transformation involves introducing regions of cloned plastid DNA flanking a selectable marker together with the gene of interest into a suitable target tissue, e.g., using biolistics or protoplast transformation (e.g., calcium chloride or PEG mediated transformation). The 1 to 1.5 kb flanking regions, termed targeting sequences, facilitate homologous recombination with the plastid genome and thus allow the replacement or modification of specific regions of the plastome. Initially, point mutations in the chloroplast 16S rRNA and rps12 genes conferring resistance to spectinomycin and/or streptomycin are utilized as selectable markers for transformation (Svab, Z., Hajdukiewicz, P., and Maliga, P. (1990) Proc. Natl. Acad. Sci. USA 87, 8526-8530; Staub, J. M., and Maliga, P. (1992) Plant Cell 4, 39-45). This resulted in stable homoplasmic transformants at a frequency of approximately one per 100 bombardments of target leaves. The presence of cloning sites between these markers allowed creation of a plastid targeting vector for introduction of foreign genes (Staub, J.M., and Maliga, P. (1993) EMBO J. 12, 601-606). Substantial increases in transformation frequency are obtained by replacement of the recessive rRNA or r-protein antibiotic resistance genes with a dominant selectable marker, the bacterial

aadA gene encoding the spectinomycin-detoxifying enzyme aminoglycoside-3'adenyltransferase (Svab, Z., and Maliga, P. (1993) Proc. Natl. Acad. Sci. USA 90, 913-917). Previously, this marker had been used successfully for high-frequency transformation of the plastid genome of the green alga Chlamydomonas reinhardtii (Goldschmidt-Clermont, M. (1991) Nucl. Acids Res. 19: 4083-4089). Other selectable markers useful for plastid transformation are known in the art and encompassed within the scope of the invention. Typically, approximately 15-20 cell division cycles following transformation are required to reach a homoplastidic state. Plastid expression, in which genes are inserted by homologous recombination into all of the several thousand copies of the circular plastid genome present in each plant cell, takes advantage of the enormous copy number advantage over nuclearexpressed genes to permit expression levels that can readily exceed 10% of the total soluble plant protein. In a preferred embodiment, a nucleotide sequence of the present invention is inserted into a plastid targeting vector and transformed into the plastid genome of a desired plant host. Plants homoplastic for plastid genomes containing a nucleotide sequence of the present invention are obtained, and are preferentially capable of high expression of the nucleotide sequence.

Formulation of Insecticidal Compositions

The invention also includes compositions comprising at least one of the insecticidal toxins of the present invention. In order to effectively control insect pests such compositions preferably contain sufficient amounts of toxin. Such amounts vary depending on the crop to be protected, on the particular pest to be targeted, and on the environmental conditions, such as humidity, temperature or type of soil. In a preferred embodiment, compositions comprising the insecticidal toxins comprise host cells expressing the toxins without additional purification. In another preferred embodiment, the cells expressing the insecticidal toxins are lyophilized prior to their use as an insecticidal agent. In another embodiment, the insecticidal toxins are engineered to be secreted from the host cells. In cases where purification of the toxins from the host cells in which they are expressed is desired, various degrees of purification of the insecticidal toxins are reached.

The present invention further embraces the preparation of compositions comprising at least one insecticidal toxin of the present invention, which is homogeneously mixed with one or more compounds or groups of compounds described herein. The present invention also relates to methods of treating plants, which comprise application of the insecticidal toxins or compositions containing the insecticidal toxins, to plants. The insecticidal toxins

can be applied to the crop area in the form of compositions or plant to be treated, simultaneously or in succession, with further compounds. These compounds can be both fertilizers or micronutrient donors or other preparations that influence plant growth. They can also be selective herbicides, insecticides, fungicides, bactericides, nematicides, molluscicides or mixtures of several of these preparations, if desired together with further carriers, surfactants or application-promoting adjuvants customarily employed in the art of formulation. Suitable carriers and adjuvants can be solid or liquid and correspond to the substances ordinarily employed in formulation technology, e.g. natural or regenerated mineral substances, solvents, dispersants, wetting agents, tackifiers, binders or fertilizers.

A preferred method of applying insecticidal toxins of the present invention is by spraying to the environment hosting the insect pest like the soil, water, or foliage of plants. The number of applications and the rate of application depend on the type and intensity of infestation by the insect pest. The insecticidal toxins can also penetrate the plant through the roots via the soil (systemic action) by impregnating the locus of the plant with a liquid composition, or by applying the compounds in solid form to the soil, e.g. in granular form (soil application). The insecticidal toxins may also be applied to seeds (coating) by impregnating the seeds either with a liquid formulation containing insecticidal toxins, or coating them with a solid formulation. In special cases, further types of application are also possible, for example, selective treatment of the plant stems or buds. The insecticidal toxins can also be provided as bait located above or below the ground.

The insecticidal toxins are used in unmodified form or, preferably, together with the adjuvants conventionally employed in the art of formulation, and are therefore formulated in known manner to emulsifiable concentrates, coatable pastes, directly sprayable or dilutable solutions, dilute emulsions, wettable powders, soluble powders, dusts, granulates, and also encapsulations, for example, in polymer substances. Like the nature of the compositions, the methods of application, such as spraying, atomizing, dusting, scattering or pouring, are chosen in accordance with the intended objectives and the prevailing circumstances.

The formulations, compositions or preparations containing the insecticidal toxins and, where appropriate, a solid or liquid adjuvant, are prepared in known manner, for example by homogeneously mixing and/or grinding the insecticidal toxins with extenders, for example solvents, solid carriers and, where appropriate, surface-active compounds (surfactants).

Suitable solvents include aromatic hydrocarbons, preferably the fractions having 8 to 12 carbon atoms, for example, xylene mixtures or substituted naphthalenes, phthalates

such as dibutyl phthalate or dioctyl phthalate, aliphatic hydrocarbons such as cyclohexane or paraffins, alcohols and glycols and their ethers and esters, such as ethanol, ethylene glycol monomethyl or monoethyl ether, ketones such as cyclohexanone, strongly polar solvents such as N-methyl-2-pyrrolidone, dimethyl sulfoxide or dimethyl formamide, as well as epoxidized vegetable oils such as epoxidized coconut oil or soybean oil or water.

The solid carriers used e.g. for dusts and dispersible powders, are normally natural mineral fillers such as calcite, talcum, kaolin, montmorillonite or attapulgite. In order to improve the physical properties it is also possible to add highly dispersed silicic acid or highly dispersed absorbent polymers. Suitable granulated adsorptive carriers are porous types, for example pumice, broken brick, sepiolite or bentonite; and suitable nonsorbent carriers are materials such as calcite or sand. In addition, a great number of pregranulated materials of inorganic or organic nature can be used, e.g. especially dolomite or pulverized plant residues.

Suitable surface-active compounds are nonionic, cationic and/or anionic surfactants having good emulsifying, dispersing and wetting properties. The term "surfactants" will also be understood as comprising mixtures of surfactants. Suitable anionic surfactants can be both water-soluble soaps and water-soluble synthetic surface-active compounds.

Suitable soaps are the alkali metal salts, alkaline earth metal salts or unsubstituted or substituted ammonium salts of higher fatty acids (chains of 10 to 22 carbon atoms), for example the sodium or potassium salts of oleic or stearic acid, or of natural fatty acid mixtures which can be obtained for example from coconut oil or tallow oil. The fatty acid methyltaurin salts may also be used.

More frequently, however, so-called synthetic surfactants are used, especially fatty sulfonates, fatty sulfates, sulfonated benzimidazole derivatives or alkylarylsulfonates.

The fatty sulfonates or sulfates are usually in the form of alkali metal salts, alkaline earth metal salts or unsubstituted or substituted ammonium salts and have a 8 to 22 carbon alkyl radical which also includes the alkyl moiety of alkyl radicals, for example, the sodium or calcium salt of lignonsulfonic acid, of dodecylsulfate or of a mixture of fatty alcohol sulfates obtained from natural fatty acids. These compounds also comprise the salts of sulfuric acid esters and sulfonic acids of fatty alcohol/ethylene oxide adducts. sulfonated benzimidazole derivatives preferably contain 2 sulfonic acid groups and one fatty acid radical containing 8 to 22 carbon atoms. Examples of alkylarylsulfonates are the sodium, calcium triethanolamine salts of dodecylbenzenesulfonic or dibutylnapthalenesulfonic acid, or of naphthalenesulfonic acid/formaldehyde а

condensation product. Also suitable are corresponding phosphates, e.g. salts of the phosphoric acid ester of an adduct of p-nonylphenol with 4 to 14 moles of ethylene oxide.

Non-ionic surfactants are preferably polyglycol ether derivatives of aliphatic or cycloaliphatic alcohols, or saturated or unsaturated fatty acids and alkylphenols, said derivatives containing 3 to 30 glycol ether groups and 8 to 20 carbon atoms in the (aliphatic) hydrocarbon moiety and 6 to 18 carbon atoms in the alkyl moiety of the alkylphenols.

Further suitable non-ionic surfactants are the water-soluble adducts of polyethylene oxide with polypropylene glycol, ethylenediamine propylene glycol and alkylpolypropylene glycol containing 1 to 10 carbon atoms in the alkyl chain, which adducts contain 20 to 250 ethylene glycol ether groups and 10 to 100 propylene glycol ether groups. These compounds usually contain 1 to 5 ethylene glycol units per propylene glycol unit.

Representative examples of non-ionic are nonylphenolpolyethoxyethanols, castor oil polyglycol ethers, polypropylene/polyethylene oxide adducts, tributylphenoxypolyethoxyethanol, polyethylene glycol and octylphenoxyethoxyethanol. Fatty acid esters of polyoxyethylene sorbitan and polyoxyethylene sorbitan trioleate are also suitable non-ionic surfactants.

Cationic surfactants are preferably quaternary ammonium salts which have, as N-substituent, at least one C8-C22 alkyl radical and, as further substituents, lower unsubstituted or halogenated alkyl, benzyl or lower hydroxyalkyl radicals. The salts are preferably in the form of halides, methylsulfates or ethylsulfates, e.g. stearyltrimethylammonium chloride or benzyldi(2-chloroethyl)ethylammonium bromide.

The surfactants customarily employed in the art of formulation are described, for example, in "McCutcheon's Detergents and Emulsifiers Annual," MC Publishing Corp. Ringwood, New Jersey, 1979, and Sisely and Wood, "Encyclopedia of Surface Active Agents," Chemical Publishing Co., Inc. New York, 1980.

EXAMPLES

The invention will be further described by reference to the following detailed examples. These examples are provided for purposes of illustration only, and are not intended to be limiting unless otherwise specified. Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by Ausubel (ed.), Current Protocols in Molecular Biology, John Wiley and Sons, Inc. (1994); T. Maniatis, E. F. Fritsch and J. Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor laboratory, Cold Spring Harbor, NY (1989); and by T.J. Silhavy, M.L. Berman, and L.W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984).

A. Isolation Of Nucleotide Sequences Whose Expression Results In Toxins Active Against Lepidopteran Insects

Example 1: Construction of Cosmid Library from Photorhabdus luminescens

Photorhabdus luminescens strain ATCC 29999 is grown in nutrient broth at 25°C for three days as described in the ATCC protocol for bioassay. The culture is grown for 24 hours for DNA isolation. Total DNA is isolated by treating freshly grown cells resuspended in 100 mM Tris pH 8, 10 mM EDTA with 2 mg/ml lysozyme for 30 minutes at 37°C. Proteinase K is added to a final concentration of 100 mg/ml, SDS is added to a final concentration of 0.5% SDS and the sample is incubated at 45°C. After the solution becomes clear and viscous, the SDS concentration is raised to 1%, and 300 mM NaCl and an equal volume of phenol-chloroform-isoamyl alcohol are added, mixed gently for 5 minutes and centrifuged at 3K. The phenol-chloroform-isoamyl alcohol extraction is repeated twice. The aqueous phase is mixed with 0.7 volumes isopropanol, and the sample is centrifuged. The pellet is washed 3 times with 70% ethanol and the nucleic acids are gently resuspended in 0.5X TE.

The DNA is treated with 0.3 units of Sau3A per mg DNA at 37°C for 3.5 minutes in 100 ml volume containing a total of 6 mg DNA. The reaction is then heated for 30 minutes at 65°C to inactivate the enzyme. Then 2 units of Calf Intestinal Alkaline Phosphatase are added and incubated for 30 minutes at 37°C. The sample is mixed with an equal volume of

phenol-chloroform-isoamyl alcohol and centrifuged. The aqueous phase is removed, precipitated with 0.7 volume isopropanol and centrifuged. The supernatant is transferred to a fresh tube, precipitated with ethanol, and the nucleic acids are resuspended in 0.5X TE at a concentration of 100 hg/ml.

SuperCos cosmid vector (Stratagene, La Jolla, CA) is prepared as described by the supplier utilizing the *BamHI* cloning site. Prepared SuperCos at 100 hg/ml is ligated with the *Sau3A* digested *P.luminescens* DNA at a molar ratio of 2:1 in a 5 ml volume overnight at 6°C. The ligation mixture is packaged using Gigapack XL III (Stratagene), as described by the supplier. Packaged phages are used to infect XL-1MR (Stratagene) cells as described by the supplier. The cosmid library is plated on L-agar with 50 mg/ml kanamycin and incubated 16 hours at 37°C. 500 colonies are patched onto fresh L-kan plates at 50 colonies per plate. From the other plates the cells are washed off with L broth and mixed with 20% glycerol and frozen at -80°C.

Example 2: Insect Bioassays

Plutella xylostella bioassays are performed by aliquoting of 50 μl of the *E. coli* culture on the solid artificial *Plutella xylostella* diet (Biever and Boldt, *Annals of Entomological Society of America*, 1971; Shelton et al., *J. Ent. Sci.* 26:17). 4 ml of the diet is poured into 1 oz. clear plastic cups (Bioserve product #9051). 5 neonate *P. xylostella* from a diet adapted lab colony are placed in each diet-containing cup and then covered with a white paper lid (Bioserve product #9049). 10 larvae are assayed per concentration. Trays of cups are placed in an incubator for 3 days at 72°F with a 14:10 (hours) light:dark cycle. Then, the number of live larvae in each cup is recorded. Bioassays for other insects are performed as described for *Plutella xylostella*, but using the diet required by the insect to be tested.

The broth of *P. luminescens* undiluted and diluted 1:100 gives 100% mortality against *P. xylostella*. The broth of *P. luminescens* also gives 100% mortality against *Diabrotica virgifera virgifera*. Three clones with activity against *P. xylostella* and *Heliothis virescens* are obtained after screening 500 *E. coli* clones by insect bioassay. These cosmid clones are given the numbers pCIB9349, pCIB9350, and pCIB9351.

- 31 -

Example 3: Isolation of the Nucl otid Sequence Responsible for Insect C ntrol Activity from Clones pClB9349, pClB9350, and pClB9351

The three clones pClB9349, pClB9350 and pClB9351 are found to be overlapping cosmids by restriction enzyme mapping. After digestion with PacI, clones pCIB9349 and pCIB9351 give two DNA fragments each, and pCIB9350 gives three DNA fragments. Each fragment is isolated and is self-ligated. The enzyme Pacl does not cut the SuperCos vector; therefore, only fragments linked to it are re-isolated. The ligation mixtures are transformed into DH5α E. coli cells. Isolated transformed bacterial colonies are grown in L broth with 50 μg/ml kanamycin, and plasmid DNA is isolated by using the alkaline miniprep protocol as described in Sambrook, et al. DNA is digested with Notl/Pacl and two clones, pCIB9355 and pCIB9356, are found by bioassay to still contain the insecticidal activity. Clone pCIB9355 is digested with Not and a 17 kb and a 4 kb DNA fragment are generated. The 17 kb fragment is isolated and ligated into Bluescript vector previously cut with Notl and transformed into DH5\alpha E. coli cells. The isolated transformed bacterial colonies are grown as described and plasmid DNA is isolated by the alkaline miniprep protocol. A clone containing the 17 kb insert is named pCIB9359 and tested by bioassay. The results are shown in Example 5. 3 µg of the 17 kb insert is isolated and treated with 0.3 unit of Sau3A per µg DNA for 4, 6, and 8 minutes at 37°C, heated at 75°C for 15 minutes. The samples are pooled and ligated into pUC19 previously cut with BamHI and treated with calf intestinal alkaline phosphatase. The ligation is transformed into DH5lpha cells and plated on L agar with Xgal/Amp as described in Sambrook et al. and grown overnight at 37°C. White colonies are picked and grown in L broth with 100 µg/ml and plasmid DNA is isolated as previously described. DNA is digested with EcoRI/HindIII and novel restriction patterns are sequenced. Sequencing primers are ordered from Genosys Biotechnologies (Woodlands, TX). Sequencing is performed using the dideoxy chain-termination method. Sequencing is completed using Applied Biosystems Inc. model 377 automated DNA sequencer (Foster City, CA). Sequence is assembled using 3.0 from Gene Codes Corporation (Ann Arbor, MI).

Exampl 4: Subcloning fth 9.7 kb EcoRI/Xbal Fragment Fr m pCIB9359

pClB9359 is digested with EcoRI and XbaI and the DNA is run on a 0.8% Seaplaque/TBE gel. The 9.7 kb fragment (SEQ ID NO:1) is isolated and ligated into pUC19 previously digested with EcoRI and XbaI. The ligation mixture is transformed into DH5 α E. coli cells. Transformed bacteria are grown and plasmid DNA is isolated as previously described. The vector containing the 9.7 kb fragment in pUC19 is designated pClB9359-7 and bioassay results are shown in Example 5.

Example 5: Bioassay Results for Cosmid Clones pCIB9359 and pCIB9359-7

Cultures of *E. coli* strains 9359 and 9359-7 containing clones pCIB9359 and pCIB9359-7, respectively, are tested for insecticidal activity against the following insects in insect bioassays:

Insects	Clones
	pClB9359 and pClB9359-7
Plutella xylostella (Diamondback Moth (DBM))	+++
Heliothis virescens (Tobacco Budworm (TBW))	++
Helicoverpa zea (Corn Earworm (CEW))	+++
Spodoptera exigua (Beet Armyworm (BAW))	+
Spodoptera frugiperda (Fall Armyworm (FAW))	+
Trichoplusia ni (Cabbage Looper (CL))	+++
Ostrinia nubilalis (European Corn Borer (ECB))	++
Manduca sexta (Tobacco Hornworm (THW)	na
Diabrotica virgifera (Western Corn Rootworm (WCR))	na
Agrotis ipsilon (Black Cutworm (BCW))	na

na = not active

- + = significant growth inhibition
- ++ = >40% mortality, but less than 100%
- +++ = 100% mortality

The clones show insecticidal activity against *P. xylostella*, *H. virescens*, *H. zea*, *T. ni*, and *O. nubilalis*, and significant insect control activity against *S. exigua* and *S. frugiperda*.

Example 6: Identification of Active Region of pClB9359-7 By Subcloning

Cultures of *E. coli* strains containing subclones of pClB9359-7 are tested for insecticidal activity in insect bioassays against *P. xylostella*.

Restriction	Nucleotide Posit	ion Relative to 9.7 kb	Insecticidal Activity Against	
Fragment	EcoRI/Xbal fragment (SEQ ID NO:1)		Plutella xylostella	
	from pCIB9539-7	and Size in kb		
EcoRI/XbaI	1 to 9712	9.7 kb	+++	
EcoRV	(-912) to 2309	3.2 kb	na	
HindIII	665 to 5438	4.7 kb	na	
Kpnl	1441 to 8137	6.9 kb	na	
Sacl/Xbal	2677 to 9712	7.0 kb	na	

na = not active

Example 7: Characterization of pCIB9359-7 Insect Control Activity By Titration

Dilutions of a culture of E.coli strain 9359-7 containing pCIB9359-7 are tested for insecticidal activity in insect bioassays. Dilutions are prepared in a culture of E.coli XL-1 in a total volume of 100 μ l and are transferred to diet cups with 5 insects per cup. The results show the percentage (%) of insect mortality.

^{+ =} significant growth inhibition

^{++ = &}gt;40% mortality, but less than 100%

^{+++ = 100%} mortality

WO 99/42589 PCT/EP99/01015

- 34 -

μ l 9359-7 Culture	Px	Hv	Hz	Tn	
100	100	72	48	100	
50	100	84	68	92	
25	100	52	32	100	
12.5	96	52	36	68	
6.25	88	20	4	32	
0	36	20	24	0	

Px = P. xylostella, Hv = H. virescens, Hz = H. zea, Tn = T. ni.

Cultures of E. coli 9359-7 still show substantial insecticidal activity after dilution.

Example 8: Stability of pCIB9359-7 Activity

The stability of the toxins is tested after storage for 2 weeks at different temperatures and conditions. 300 ml of Luria broth containing 100 (µg/ml ampicillin is inoculated with *E. coli* strain 9359-7 and grown overnight at 37°C. Samples are placed in sterile 15 ml screw cap tubes and stored at 22°C and 4°C. Another sample is centrifuged; the supernatant is removed, freeze dried and stored at 22°C. The samples are stored under these conditions for 2 weeks and then a bioassay is conducted against *P. xylostella*. The freeze dried material is resuspended in the same volume as before. All samples are resuspended by vortexing.

Conditions	Results	
22°C (2 weeks)	+++	
4°C (2 weeks)	+++	:
Freeze Dried (2 weeks)	+++	

na = not active; + = significant growth inhibition; ++ = >40% mortality, but less than 100%; +++ = 100% mortality

WO 99/42589 PCT/EP99/01015

- 35 -

This demonstrates that the toxins retain their activity for at least two weeks at 22°C, 4°C, and freeze-dried, and are therefore very stable.

Example 9: Size Fraction of pCIB9359-7 Activity

The approximate sizes of the insecticidal toxins are determined. P. luminescens cosmid clones pClB9359-7 and pUC19 in E. coli host DH5α are grown in media consisting of 50% Terrific broth and 50% Luria broth, supplemented with 50 μg/ml ampicillin. Cultures (three tubes of each strain) are inoculated into 3 ml of the above media in culture tubes and incubated on a roller wheel overnight at 37°C. Cultures of each strain are combined and sonicated using a Branson Model 450 Sonicator, micro tip, for approximately six 10 second cycles with cooling on ice between cycles. The sonicates are centrifuged in a Sorvall SS34 rotor at 6000 RPM for 10 minutes. The resultant supernatants are filtered through a 0.2 µ filter. The 3 ml fractions of the filtrates are applied to Bio-Rad Econo-Pac 10DG columns that have been previously equilibrated with 10 ml of 50mM NaCl, 25 mM Tris base, pH 7.0. The flow through collected during sample loading is discarded. The samples are fractionated with two subsequent additions of 4 ml each of the NaCl - Tris equilibration buffer. The two four ml fractions are saved for testing. The first fraction contains all material above about 6,000 mol. wt; the second fraction contains material smaller than 6,000 mol. wt. A sample of the whole culture broth, the sonicate, and the filtered supernatant on the sonicate are tested along with the three fractions from the 10DG column for activity on P. xylostella neonates in bioassays.

The culture, the sonicate, and the filtered supernatant of the sonicate, and the first column fraction from the 9359-7 sample are highly active on *P. xylostella*. The second column fraction from 9359-7 is slightly active (some stunting only). No activity is found in the third fraction from 9359-7. The sample from DH5-pUC19 does not have any activity. This indicates that the molecular weights of the toxins are above 6,000.

Example 10: Heat Inactivitation of pCIB9359-7 Activity

The heat stability of the toxins is determined. Overnight cultures of the *E. coli* strain pClB9359-7 are grown in a 50:50 mixture of Luria broth and Terrific broth. Cultures are grown at 37°C in culture tubes on a tube roller. A one ml sample of the culture is placed in

a 1.5 ml eppendorf tube and placed in a boiling water bath. The sample is removed after five minutes and allowed to cool to room temperature. This sample along with an untreated portion of the culture is assayed on *P. xylostella*. 50µl of sample of sample is spread on diet, allowed to dry and neonate larvae *P. xylostella* applied to the surface. The assay is incubated for 5 days at room temperature.

The untreated sample causes 100% mortality. The heat treated sample and a diet alone control do not cause any observable mortality, showing the toxins are heat sensitive.

Example 11: Leaf Dip Bioassay of pCIB9359-7

Insecticidal activity of the toxins is tested in a leaf dip bioassay. Six leaves approximately 2cm in diameter each are cut from seedlings of turnip and placed in a 1oz. plastic cup (Jet Plastica) with 4ml-5ml of the resuspended toxin, covered tightly, and shaken until thoroughly wetted. The treated leaves are placed in 50mm petri dishes (Gelman Sciences) on absorbent pads moistened with 300µl of water. The dish covers are left open until the leaf surface appears dry and then placed on tightly so that the leaves do not dry out.

Ten neonate *P. xylostella* larvae are placed in each petri dish arena. Also, a treatment of 0.1% Bond spreader/sticker with no toxin is set up as a control. The arenas are monitored daily for signs of drying leaves, and water is added or leaves replaced if necessary. After 3 days the leaves and arenas are examined under a dissecting microscope, and the number of live larvae in each arena is recorded.

100% mortality is found for 9359-7 and none in the no-toxin control, showing that the toxins are also insecticidal in a leaf dip assay.

B. Isolation Of Nucleic Acid Sequences Whose Expression Results In Toxins Active Against Lepidopteran and Coleopteran Insects

Example 12: Total DNA Isolation from Photorhabdus luminescens

Photorhabdus luminescens strain ATCC 29999 is grown 14-18 hours in L broth. Total DNA is isolated from 1.5 mls of culture resuspended in 0.5% SDS, 100μg/ml proteinase K, TE to a final volume of 600 μl. After a 1 hour incubation at 37°C, 100μl 5M

NaCl and 80µl CTAB/NaCl are added and the culture is incubated at 65°C for 10 minutes. An equal volume of chloroform is added; the culture is mixed gently and spun. The aqueous phase is extracted once with phenol and once with chloroform. The nucleic acids are treated with 10 µg RNase A for 30 minutes at room temperature. The aqueous phase is mixed with 0.6 volumes isopropanol and the sample is centrifuged. The pellet is washed once with 70% ethanol and the nucleic acids are gently resuspended in 100-200ul TE.

Example 13: PCR Amplification of Probes

Two probes are PCR amplified from *Photorhabdus luminescens* strain ATCC 29999 genomic DNA using oligos 5'-ACACAGCAGGTTCGTCAG-3' (SEQ ID NO:7) and 5'-GGCAGAAGCACTCAACTC-3' (SEQ ID NO:8) to amplify probe #1 and oligos 5'-ATTGATAGCACGCGGCGACC-3' (SEQ ID NO:9) and 5'-

TTGTAACGTGGAGCCGAACTGG-3' (SEQ ID NO:10) to amplify probe #2. The oligos are ordered from Genosys Biotechnologies, Inc. (Texas). Approximately 10-50 ng of genomic DNA is used as the template. 0.8μM of oligos, 200μM of dNTPs, 1X Taq DNA Polymerase buffer and 2.5 units of Taq DNA Polymerase are included in the reaction. The reaction conditions are as follows:

94°C - 1 minute

94°C - 30 seconds / 60°C - 30 seconds / 72°C - 30 seconds (25 cycles)

72°C - 5 minutes

4°C - indefinite soak

The reactions are preferably carried out in a PCR System 9600 (Perkin Elmer) thermocycler.

Example 14: Probing a Photorhabdus luminescens Library

600 clones from the *P. luminescens* cosmid library described in Example 1 are patched to L-amp plates in duplicate. The colonies are grown overnight then moved to 4°C. The colonies are lifted onto Colony/Plaque Screen Hybridization Transfer Membranes (Biotechnology Systems NEN Research Products). The membranes are incubated 2-3 minutes in 0.75ml 0.5N NaOH twice. The membranes are then incubated 2-3 minutes in

0.75ml 1.0M Tris-HCl, pH 7.5 twice. The membranes are allowed to dry at room temperature.

Probe #1 and probe #2 described in Example 13 are labeled using the DECAprime II Kit as described by the manufacturer (Ambion cat# 1455). Unincorporated nucleotides are removed from the labeled probes using Quick Spin Columns as described by the manufacturer (Boehringer Mannheim cat #1273973). The labeled probes are measured for incorporated radioactivity and the specific activity is 10,000,000 cpm. Membranes are prewetted with 2X SSC and hybridized with the probes for 12-16 hours at 65°C. One set of colony lifts is hybridized with probe #1 and the other set is hybridized with probe #2. The membranes are washed with wash CHURCH solutions 1 and 2 (Church and Gilbert, *Proc. Natl. Acad. Sci. USA* 81:1991-1995 (1984)) and exposed to Kodak film.

Twenty one clones are identified that hybridize to probe #1 and seven clones are identified that hybridize to probe #2. The gene in the clones isolated with probe #1 is named *hph1* and the gene in the clones isolated with probe #2 is named *hph2*.

Example 15: Insect Bioassays

The clones identified in Example 14 are tested for insecticidal activity against the following insects in insect bioassays: *Diabrotica virgifera virgifera* (Western Corn Rootworm (WCR)), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm (SCR)), *Ostrinia nubilalis* (European Corn Borer (ECB)), and *Plutella xylostella* (Diamondback Moth (DBM)).

Diabrotica virgifera virgifera (Western Corn Rootworm) and Diabrotica undecimpunctata howardi (Southern Corn Rootworm) assays are performed using a diet incorporation method. 500μl of an overnight culture of the cosmid library in XL-1 Blue MR cells (Stratagene) is sonicated and then mixed with 500μl of diet. Once the diet solidifies, it is dispensed in a petri dish and 20 larvae are introduced over the diet. Trays of dishes are placed in an incubator for 3-5 days, and percent mortality is recorded at the end of the assay period.

Ostrinia nubilalis (European Corn Borer) and Plutella xylostella (Diamondback Moth) assays are performed by a surface treatment method. The diet is poured in the petri dish and allowed it to solidify. The E. coli culture of 200 -300µl volume is dispensed over the diet surface and entire diet surface is covered to spread the culture with the help of bacterial loop. Once the surface is dry, 10 larvae are introduced over the diet surface. Trays of

dishes are placed in an incubator for 3-5 days. The assay with European Corn Borer is incubated at 30°C in complete darkness; the assay with Diamondback Moth is incubated at 72°F with a 14:10 (hours) light:dark cycle. Percent mortality is recorded at the end of the assay period.

Cosmids containing *hph2* are identified with a range of activities, including: WCR only; SCR only; WCR and SCR; SCR and ECB; WCR, SCR, and ECB; or WCR, SCR, ECB, and DBM activity.

In addition to probing the *P. luminescens* cosmid library with DNA probes, 600 clones are screened by Western Corn Rootworm bioassay. A clone is identified with activity against Western Corn Rootworm. This clone hybridizes with probe #2.

From these bioassays, cosmid 514, having activity against WCR, SCR, ECB, and DBM, is selected for sequencing.

Example 16: Sequencing of Cosmid 514

Cosmid 514 is sequenced using dye terminator chemistry on an ABI 377 instrument. The nucleotide sequence of cosmid 514 is set forth as SEQ ID NO:11. Cosmid 514 is designated pNOV2400 and deposited with the NRRL in $E.\ coli\ DH5\alpha$ and assigned accession no. B-30077.

Example 17: Subcloning Insecticidal Regions of Cosmid 514

514a

An 9011 base pair fragment within cosmid 514 (SEQ ID NO:11) is removed by digesting the cosmid with the restriction endonuclease *Spel* (New England Biolabs (Massachusetts), and ligating (T4 DNA Ligase, NEB) the remainder of 514. Subclone 514a consists of cosmid 514 DNA from base pairs 1-2157 ligated to base pairs 11,169-37,948.

H2O2/pET34

hph2 and orf2 (SEQ ID NO:11, base pairs 23,768-35,838) are cloned into pET34b (Novagen, Wisconsin). Restriction sites are engineered on both ends of each gene to facilitate cloning. PCR is used to add the restriction sites to the genes. A BamHI site is on the 5' end of hph2 immediately upstream of the ATG of hph2, and a Sac site is added to

the 3' end of hph2 immediately following the DNA triplet encoding the stop codon. A guanidine is added between the BamHI site and the start codon of hph2 to put the hph2 gene in frame with the Cellulose Binding Domain tag in pET34b. Orf2 has a SacI site upstream of the 56 base pairs between the stop codon of hph2 and the start codon of orf2. The 56 base pairs are included in the hph2-orf2 construct to mimic their setup in the 514 cosmid. Orf2 has an XhoI site on the 3' end immediately following the stop codon. The oligos used to add the restriction sites to hph2 and orf2 are as follows:

hph2-A	5'-CGGGATCCGATGATTTTAAAAGG-3' (SEQ ID NO:15)
hph2-B	5'-GCGCCATTGATTTGAG-3' (SEQ ID NO:16)
hph2-C	5'-CATTAGAGGTCGAACGTAC-3' (SEQ ID NO:17)
hph2-D	5'-GAGCGAGCTCTTACTTAATGGTGTAG-3' (SEQ ID NO:18)
orf2-A3	5'-CAGCGAGCTCCATGCAGAATTCACAGAC-3' (SEQ ID NO:19)
orf2-B	5'-GGCAATGGCAGCGATAAG-3' (SEQ ID NO:20)
orf2-C	5'-CATTAACGCAGGAAGAGC-3' (SEQ ID NO:21)
orf2-D	5'-GACCTCGAGTTACACGAGCGCGTCAG-3' (SEQ ID NO:22)

The BamHI-Sacl 7583 base pair fragment, corresponding to the hph2 gene, and the Sacl-Xhol 4502 base pair orf2 (including the 56 base pairs between hph2 and orf2 open reading frames), corresponding to orf2, are ligated with BamHI-Xhol-digested vector DNA pET34b.

Orf5/pBS (Noti-BamHI)

The 5325 base pair *Not*I-*Bam*HI fragment of cosmid 514 is cloned into pBS-SK using *Aff*III-*Not*I (415 bp) and *Bam*HI-*Aff*III (2530 bp) fragments of pBS-SK.

O5-H2-O2

The 12,031 base pair *BamHI-XhoI* fragment of H2O2/pET34 is cloned into the 8220 base pair *XhoI-BamHI* fragment of Orf5/pBS.

O51011H2O2

A 7298 base pair *Bam*HI-*Mlu*I fragment from subclone 514a is ligated (T4 DNA Ligase, NEB) with 9588 bp *Mlu*I-*Xho*I and 8220 bp *Xho*I-*Bam*HI fragments of subclone O5-H2-O2. The resulting ~ 22 kb subclone O51011H2O2, which has activity against WCR and

ECB, is designated pNOV1001 and deposited with the NRRL in *E. coli* DH5 α and assigned accession no. B-30078.

AKH2O2

A 12,074 base pair *BamHI-Avr*II fragment of H2O2/pET34 is ligated (T4 DNA Ligase, NEB) into pK184 *Nhel-BamHI* fragment (2228 bp), generating a clone containing hph2 and orf2 in a p15a origin of replication, kanamycin-resistant vector.

Example 18: Insecticidal Activity of Subclones

Bioassays as described above are performed with *E. coli* cultures that express the above subclones, both singly and in combination. Coexpressing AKH2O2 and Orf5/pBS in *E. coli*, for example in DH5α or HB101, is found to give insecticidal activity against the Lepidopterans *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), as well as against the Coleopterans *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle). Thus, coexpression of hph2 (SEQ ID NO:11, base pairs 23,768-31,336), orf2 (SEQ ID NO:11, base pairs 31,393-35,838), and orf5 (SEQ ID NO:11, base pairs 15,171-18,035) is sufficient to control these insects. In addition, expression of each of these three ORFs on separate plasmids gives insect control activity, demonstrating that they do not have to be genetically linked to be active, so long as all three gene products are present.

C. Expression of the Nucleic Acid Sequences of the Invention in Heterologous Microbial Hosts

Microorganisms which are suitable for the heterologous expression of the nucleotide sequences of the invention are all microorganisms which are capable of colonizing plants or the rhizosphere. As such they will be brought into contact with insect pests. These include gram-negative microorganisms such as *Pseudomonas*, *Enterobacter* and *Serratia*, the gram-positive microorganism *Bacillus* and the fungi *Trichoderma*, *Gliocladium*, and *Saccharomyces cerevisiae*. Particularly preferred heterologous hosts are *Pseudomonas fluorescens*, *Pseudomonas putida*, *Pseudomonas cepacia*, *Pseudomonas aureofaciens*,

Pseudomonas aurantiaca, Enterobacter cloacae, Serratia marscesens, Bacillus subtilis, Bacillus cereus, Trichoderma viride, Trichoderma harzianum, Gliocladium virens, and Saccharomyces cerevisiae.

Example 19: Expression of the Nucleotide Sequences in *E. coli* and Other Gram-Negative Bacteria

Many genes have been expressed in gram-negative bacteria in a heterologous manner. Expression vector pKK223-3 (Pharmacia catalogue # 27-4935-01) allows expression in *E. coli*. This vector has a strong *tac* promoter (Brosius, J. *et al.*, *Proc. Natl. Acad. Sci. USA 81*) regulated by the *lac* repressor and induced by IPTG. A number of other expression systems have been developed for use in *E. coli*. The thermoinducible expression vector pPL (Pharmacia #27-4946-01) uses a tightly regulated bacteriophage λ promoter which allows for high level expression of proteins. The *lac* promoter provides another means of expression but the promoter is not expressed at such high levels as the *tac* promoter. With the addition of broad host range replicons to some of these expression system vectors, expression of the nucleotide sequence in closely related gram negative-bacteria such as *Pseudomonas*, *Enterobacter*, *Serratia* and *Erwinia* is possible. For example, pLRKD211 (Kaiser & Kroos, Proc. Natl. Acad. Sci. USA <u>81</u>: 5816-5820 (1984)) contains the broad host range replicon *ori T* which allows replication in many gram-negative bacteria.

In *E. coli*, induction by IPTG is required for expression of the *tac* (*i.e. trp-lac*) promoter. When this same promoter (*e.g.* on wide-host range plasmid pLRKD211) is introduced into *Pseudomonas* it is constitutively active without induction by IPTG. This *trp-lac* promoter can be placed in front of any gene or operon of interest for expression in *Pseudomonas* or any other closely related bacterium for the purposes of the constitutive expression of such a gene. Thus, a nucleotide sequence whose expression results in an insecticidal toxin can therefore be placed behind a strong constitutive promoter, transferred to a bacterium which has plant or rhizosphere colonizing properties turning this organism to an insecticidal agent. Other possible promoters can be used for the constitutive expression of the nucleotide sequence in gram-negative bacteria. These include, for example, the promoter from the *Pseudomonas* regulatory genes *gafA* and *lemA* (WO 94/01561) and the

Pseudomonas savastanoi IAA operon promoter (Gaffney et al., J. Bacteriol. 172: 5593-5601 (1990).

Example 20: Expression of the Nucleotide Sequences in Gram-Positive Bacteria

Heterologous expression of the nucleotides sequence in gram-positive bacteria is another means of producing the insecticidal toxins. Expression systems for *Bacillus* and *Streptomyces* are the best characterized. The promoter for the erythromycin resistance gene (*ermR*) from *Streptococcus pneumoniae* has been shown to be active in gram-positive aerobes and anaerobes and also in *E.coli* (Trieu-Cuot *et al.*, Nucl Acids Res 18: 3660 (1990)). A further antibiotic resistance promoter from the thiostreptone gene has been used in *Streptomyces* cloning vectors (Bibb, Mol Gen Genet 199: 26-36 (1985)). The shuttle vector pHT3101 is also appropriate for expression in *Bacillus* (Lereclus, FEMS Microbiol Lett 60: 211-218 (1989)). A significant advantage of this approach is that many grampositive bacteria produce spores which can be used in formulations that produce insecticidal agents with a longer shelf life. *Bacillus* and *Streptomyces* species are aggressive colonizers of soils

Example 21: Expression of the Nucleotide Sequences in Fungi

Trichoderma harzianum and Gliocladium virens have been shown to provide varying levels of biocontrol in the field (US 5,165,928 and US 4,996,157, both to Cornell Research Foundation). A nucleotide sequence whose expression results in an insecticidal toxin could be expressed in such a fungus. This could be accomplished by a number of ways which are well known in the art. One is protoplast-mediated transformation of the fungus by PEG or electroporation-mediated techniques. Alternatively, particle bombardment can be used to transform protoplasts or other fungal cells with the ability to develop into regenerated mature structures. The vector pAN7-1, originally developed for Aspergillus transformation and now used widely for fungal transformation (Curragh et al., Mycol. Res. 97(3): 313-317 (1992); Tooley et al., Curr. Genet. 21: 55-60 (1992); Punt et al., Gene 56: 117-124 (1987)) is engineered to contain the nucleotide sequence. This plasmid contains the E. coli the hygromycin B resistance gene flanked by the Aspergillus nidulans gpd promoter and the trpC terminator (Punt et al., Gene 56: 117-124 (1987)).

- 44 -

In a preferred embodiment, the nucleic acid sequences of the invention are expressed in the yeast *Saccharomyces cerevisiae*. Each of the three ORF's of SEQ ID NO:11 (hph2, orf2 and orf5), which together confer insecticidal activity, are cloned into individual vectors with the GAL1 inducible promoter and the CYC1 terminator. Each vector has ampicillin resistance and the 2 micron replicon. The vectors differ in their yeast growth markers. hph2 is cloned into p424 (TRP1, ATCC 87329), orf2 into p423 (HIS3, ATCC 87327), and orf5 into p425 (LEU2, ATCC 87331). The three constructs are transformed into *S. cerevisiae* independently and together. The three ORFs are expressed together and tested for protein expression and insecticidal activity.

D. Expression of the Nucleotide Sequences in Transgenic Plants

The nucleic acid sequences described in this application can be incorporated into plant cells using conventional recombinant DNA technology. Generally, this involves inserting a coding sequence of the invention into an expression system to which the coding sequence is heterologous (i.e., not normally present) using standard cloning procedures known in the art. The vector contains the necessary elements for the transcription and translation of the inserted protein-coding sequences. A large number of vector systems known in the art can be used, such as plasmids, bacteriophage viruses and other modified viruses. Suitable vectors include, but are not limited to, viral vectors such as lambda vector systems λgtl1, λgtl0 and Charon 4; plasmid vectors such as pBI121, pBR322, pACYC177, pACYC184, pAR series, pKK223-3, pUC8, pUC9, pUC18, pUC19, pLG339, pRK290, pKC37, pKC101, pCDNAII; and other similar systems. The components of the expression system may also be modified to increase expression. For example, truncated sequences, nucleotide substitutions or other modifications may be employed. The expression systems described herein can be used to transform virtually any crop plant cell under suitable Transformed cells can be regenerated into whole plants such that the conditions. nucleotide sequence of the invention confer insect resistance to the transgenic plants.

Example 22: Modification of Coding Sequences and Adjacent Sequences

The nucleotide sequences described in this application can be modified for expression in transgenic plant hosts. A host plant expressing the nucleotide sequences and

WO 99/42589 PCT/EP99/01015

- 45 -

which produces the insecticidal toxins in its cells has enhanced resistance to insect attack and is thus better equipped to withstand crop losses associated with such attack.

The transgenic expression in plants of genes derived from microbial sources may require the modification of those genes to achieve and optimize their expression in plants. In particular, bacterial ORFs which encode separate enzymes but which are encoded by the same transcript in the native microbe are best expressed in plants on separate transcripts. To achieve this, each microbial ORF is isolated individually and cloned within a cassette which provides a plant promoter sequence at the 5' end of the ORF and a plant transcriptional terminator at the 3' end of the ORF. The isolated ORF sequence preferably includes the initiating ATG codon and the terminating STOP codon but may include additional sequence beyond the initiating ATG and the STOP codon. In addition, the ORF may be truncated, but still retain the required activity; for particularly long ORFs, truncated versions which retain activity may be preferable for expression in transgenic organisms. By "plant promoter" and "plant transcriptional terminator" it is intended to mean promoters and transcriptional terminators which operate within plant cells. This includes promoters and transcription terminators which may be derived from non-plant sources such as viruses (an example is the Cauliflower Mosaic Virus).

In some cases, modification to the ORF coding sequences and adjacent sequence is not required. It is sufficient to isolate a fragment containing the ORF of interest and to insert it downstream of a plant promoter. For example, Gaffney *et al.* (Science 261: 754-756 (1993)) have expressed the *Pseudomonas nahG* gene in transgenic plants under the control of the CaMV 35S promoter and the CaMV *tml* terminator successfully without modification of the coding sequence and with x bp of the *Pseudomonas* gene upstream of the ATG still attached, and y bp downstream of the STOP codon still attached to the *nahG* ORF. Preferably as little adjacent microbial sequence should be left attached upstream of the ATG and downstream of the STOP codon. In practice, such construction may depend on the availability of restriction sites.

In other cases, the expression of genes derived from microbial sources may provide problems in expression. These problems have been well characterized in the art and are particularly common with genes derived from certain sources such as *Bacillus*. These problems may apply to the nucleotide sequence of this invention and the modification of these genes can be undertaken using techniques now well known in the art. The following problems may be encountered:

WO 99/42589 PCT/EP99/01015

- 46 -

1. Codon Usage.

The preferred codon usage in plants differs from the preferred codon usage in certain microorganisms. Comparison of the usage of codons within a cloned microbial ORF to usage in plant genes (and in particular genes from the target plant) will enable an identification of the codons within the ORF which should preferably be changed. Typically plant evolution has tended towards a strong preference of the nucleotides C and G in the third base position of monocotyledons, whereas dicotyledons often use the nucleotides A or T at this position. By modifying a gene to incorporate preferred codon usage for a particular target transgenic species, many of the problems described below for GC/AT content and illegitimate splicing will be overcome.

2. GC/AT Content.

Plant genes typically have a GC content of more than 35%. ORF sequences which are rich in A and T nucleotides can cause several problems in plants. Firstly, motifs of ATTTA are believed to cause destabilization of messages and are found at the 3' end of many short-lived mRNAs. Secondly, the occurrence of polyadenylation signals such as AATAAA at inappropriate positions within the message is believed to cause premature truncation of transcription. In addition, monocotyledons may recognize AT-rich sequences as splice sites (see below).

3. Sequences Adjacent to the Initiating Methionine.

Plants differ from microorganisms in that their messages do not possess a defined ribosome binding site. Rather, it is believed that ribosomes attach to the 5' end of the message and scan for the first available ATG at which to start translation. Nevertheless, it is believed that there is a preference for certain nucleotides adjacent to the ATG and that expression of microbial genes can be enhanced by the inclusion of a eukaryotic consensus translation initiator at the ATG. Clontech (1993/1994 catalog, page 210, incorporated herein by reference) have suggested one sequence as a consensus translation initiator for the expression of the *E. coli uidA* gene in plants. Further, Joshi (NAR 15: 6643-6653 (1987), incorporated herein by reference) has compared many plant sequences adjacent to the ATG and suggests another consensus sequence. In situations where difficulties are encountered in the expression of microbial ORFs in plants, inclusion of one of these sequences at the initiating ATG may improve translation. In such cases the last three

- 47 -

nucleotides of the consensus may not be appropriate for inclusion in the modified sequence due to their modification of the second AA residue. Preferred sequences adjacent to the initiating methionine may differ between different plant species. A survey of 14 maize genes located in the GenBank database provided the following results:

Position Before the Initiating ATG in 14 Maize Genes:	Position	Before	the	Initiating	ATG i	n 14	Maize	Genes:
-------------------------------------------------------	----------	--------	-----	------------	-------	------	-------	--------

	<u>-10</u>	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
C	3	8	4	6	2	5	6	0	10	7
Т	3	0	3	4	3	2	1	1	1	0
Α	2	3	1	4	3	2	3	7	2	3
G	6	3	6	0	6	5	4	6	1	5

This analysis can be done for the desired plant species into which the nucleotide sequence is being incorporated, and the sequence adjacent to the ATG modified to incorporate the preferred nucleotides.

4. Removal of Illegitimate Splice Sites.

Genes cloned from non-plant sources and not optimized for expression in plants may also contain motifs which may be recognized in plants as 5' or 3' splice sites, and be cleaved, thus generating truncated or deleted messages. These sites can be removed using the techniques well known in the art.

Techniques for the modification of coding sequences and adjacent sequences are well known in the art. In cases where the initial expression of a microbial ORF is low and it is deemed appropriate to make alterations to the sequence as described above, then the construction of synthetic genes can be accomplished according to methods well known in the art. These are, for example, described in the published patent disclosures EP 0 385 962 (to Monsanto), EP 0 359 472 (to Lubrizol) and WO 93/07278 (to Ciba-Geigy), all of which are incorporated herein by reference. In most cases it is preferable to assay the expression of gene constructions using transient assay protocols (which are well known in the art) prior to their transfer to transgenic plants.

Example 23: Construction of Plant Expr ssion Cass tt s

Coding sequences intended for expression in transgenic plants are first assembled in expression cassettes behind a suitable promoter expressible in plants. The expression cassettes may also comprise any further sequences required or selected for the expression of the transgene. Such sequences include, but are not restricted to, transcription terminators, extraneous sequences to enhance expression such as introns, vital sequences, and sequences intended for the targeting of the gene product to specific organelles and cell compartments. These expression cassettes can then be easily transferred to the plant transformation vectors described below. The following is a description of various components of typical expression cassettes.

1. Promoters

The selection of the promoter used in expression cassettes will determine the spatial and temporal expression pattern of the transgene in the transgenic plant. Selected promoters will express transgenes in specific cell types (such as leaf epidermal cells, mesophyll cells, root cortex cells) or in specific tissues or organs (roots, leaves or flowers, for example) and the selection will reflect the desired location of accumulation of the gene product. Alternatively, the selected promoter may drive expression of the gene under various inducing conditions. Promoters vary in their strength, i.e., ability to promote transcription. Depending upon the host cell system utilized, any one of a number of suitable promoters can be used, including the gene's native promoter. The following are non-limiting examples of promoters that may be used in expression cassettes.

a. Constitutive Expression, the Ubiquitin Promoter:

Ubiquitin is a gene product known to accumulate in many cell types and its promoter has been cloned from several species for use in transgenic plants (e.g. sunflower - Binet et al. Plant Science 79: 87-94 (1991); maize - Christensen et al. Plant Molec. Biol. 12: 619-632 (1989); and Arabidopsis - Norris et al., Plant Mol. Biol. 21:895-906 (1993)). The maize ubiquitin promoter has been developed in transgenic monocot systems and its sequence and vectors constructed for monocot transformation are disclosed in the patent publication EP 0 342 926 (to Lubrizol) which is herein incorporated by reference. Taylor et al. (Plant Cell Rep. 12: 491-495 (1993)) describe a vector (pAHC25) that comprises the maize ubiquitin promoter and first intron and its high activity in cell suspensions of numerous

- 49 -

PCT/EP99/01015

monocotyledons when introduced via microprojectile bombardment. The *Arabidopsis* ubiquitin promoter is ideal for use with the nucleotide sequences of the present invention. The ubiquitin promoter is suitable for gene expression in transgenic plants, both monocotyledons and dicotyledons. Suitable vectors are derivatives of pAHC25 or any of the transformation vectors described in this application, modified by the introduction of the appropriate ubiquitin promoter and/or intron sequences.

b. Constitutive Expression, the CaMV 35S Promoter:

Construction of the plasmid pCGN1761 is described in the published patent application EP 0 392 225 (Example 23), which is hereby incorporated by reference. pCGN1761 contains the "double" CaMV 35S promoter and the tml transcriptional terminator with a unique EcoRI site between the promoter and the terminator and has a pUC-type backbone. A derivative of pCGN1761 is constructed which has a modified polylinker which includes NotI and XhoI sites in addition to the existing EcoRI site. This derivative is designated pCGN1761ENX. pCGN1761ENX is useful for the cloning of cDNA sequences or coding sequences (including microbial ORF sequences) within its polylinker for the purpose of their expression under the control of the 35S promoter in transgenic plants. The entire 35S promoter-coding sequence-tml terminator cassette of such a construction can be excised by HindIII, Sphl, Sall, and Xbal sites 5' to the promoter and Xbal, BamHI and Ball sites 3' to the terminator for transfer to transformation vectors such as those described below. Furthermore, the double 35S promoter fragment can be removed by 5' excision with HindIII, SphI, Sall, Xbal, or PstI, and 3' excision with any of the polylinker restriction sites (EcoRI, NotI or XhoI) for replacement with another promoter. If desired, modifications around the cloning sites can be made by the introduction of sequences that may enhance translation. This is particularly useful when overexpression is desired. For example, pCGN1761ENX may be modified by optimization of the translational initiation site as described in Example 37 of U.S. Patent No. 5,639,949, incorporated herein by reference.

c. Constitutive Expression, the Actin Promoter:

Several isoforms of actin are known to be expressed in most cell types and consequently the actin promoter is a good choice for a constitutive promoter. In particular, the promoter from the rice *Actl* gene has been cloned and characterized (McElroy *et al.* Plant Cell 2: 163-171 (1990)). A 1.3kb fragment of the promoter was found to contain all

the regulatory elements required for expression in rice protoplasts. Furthermore, numerous expression vectors based on the ActI promoter have been constructed specifically for use in monocotyledons (McEiroy et al. Mol. Gen. Genet. 231: 150-160 (1991)). These incorporate the ActI-intron 1, AdhI 5' flanking sequence and AdhI-intron 1 (from the maize alcohol dehydrogenase gene) and sequence from the CaMV 35S promoter. Vectors showing highest expression were fusions of 35S and Actl intron or the Actl 5' flanking sequence and the ActI intron. Optimization of sequences around the initiating ATG (of the GUS reporter gene) also enhanced expression. The promoter expression cassettes described by McElroy et al. (Mol. Gen. Genet. 231: 150-160 (1991)) can be easily modified for gene expression and are particularly suitable for use in monocotyledonous hosts. For example, promotercontaining fragments is removed from the McElroy constructions and used to replace the double 35S promoter in pCGN1761ENX, which is then available for the insertion of specific gene sequences. The fusion genes thus constructed can then be transferred to appropriate transformation vectors. In a separate report, the rice Actl promoter with its first intron has also been found to direct high expression in cultured barley cells (Chibbar et al. Plant Cell Rep. <u>12</u>: 506-509 (1993)).

d. Inducible Expression, the PR-1 Promoter:

The double 35S promoter in pCGN1761ENX may be replaced with any other promoter of choice that will result in suitably high expression levels. By way of example, one of the chemically regulatable promoters described in U.S. Patent No. 5,614,395 may replace the double 35S promoter. The promoter of choice is preferably excised from its source by restriction enzymes, but can alternatively be PCR-amplified using primers that carry appropriate terminal restriction sites. Should PCR-amplification be undertaken, then the promoter should be re-sequenced to check for amplification errors after the cloning of the amplified promoter in the target vector. The chemically/pathogen regulatable tobacco PR-1a promoter is cleaved from plasmid pCIB1004 (for construction, see example 21 of EP 0 332 104, which is hereby incorporated by reference) and transferred to plasmid pCGN1761ENX (Uknes et al., 1992). pCIB1004 is cleaved with *Ncol* and the resultant 3' overhang of the linearized fragment is rendered blunt by treatment with T4 DNA polymerase. The fragment is then cleaved with *HindIII* and the resultant PR-1a promotercontaining fragment is gel purified and cloned into pCGN1761ENX from which the double 35S promoter has been removed. This is done by cleavage with *Xhol* and blunting with T4

PCT/EP99/01015

- 51 -

polymerase, followed by cleavage with HindIII and isolation of the larger vector-terminator containing fragment into which the pClB1004 promoter fragment is cloned. This generates a pCGN1761ENX derivative with the PR-1a promoter and the tml terminator and an intervening polylinker with unique EcoRl and Notl sites. The selected coding sequence can be inserted into this vector, and the fusion products (i.e. promoter-gene-terminator) can subsequently be transferred to any selected transformation vector, including those described infra. Various chemical regulators may be employed to induce expression of the selected coding sequence in the plants transformed according to the present invention, including the benzothiadiazole, isonicotinic acid, and salicylic acid compounds disclosed in U.S. Patent Nos. 5,523,311 and 5,614,395.

e. Inducible Expression, an Ethanol-Inducible Promoter:

A promoter inducible by certain alcohols or ketones, such as ethanol, may also be used to confer inducible expression of a coding sequence of the present invention. Such a promoter is for example the alcA gene promoter from Aspergillus nidulans (Caddick et al. (1998) Nat. Biotechnol 16:177-180). In A. nidulans, the alcA gene encodes alcohol dehydrogenase I, the expression of which is regulated by the AlcR transcription factors in presence of the chemical inducer. For the purposes of the present invention, the CAT coding sequences in plasmid palcA:CAT comprising a alcA gene promoter sequence fused to a minimal 35S promoter (Caddick et al. (1998) Nat. Biotechnol 16:177-180) are replaced by a coding sequence of the present invention to form an expression cassette having the coding sequence under the control of the alcA gene promoter. This is carried out using methods well known in the art.

f. Inducible Expression, a Glucocorticoid-Inducible Promoter:

Induction of expression of a nucleic acid sequence of the present invention using systems based on steroid hormones is also contemplated. For example, a glucocorticoidmediated induction system is used (Aoyama and Chua (1997) The Plant Journal 11: 605-612) and gene expression is induced by application of a glucocorticoid, for example a synthetic glucocorticoid, preferably dexamethasone, preferably at a concentration ranging from 0.1mM to 1mM, more preferably from 10mM to 100mM. For the purposes of the present invention, the luciferase gene sequences are replaced by a nucleic acid sequence of the invention to form an expression cassette having a nucleic acid sequence of the WO 99/42589 PCT/EP99/01015

- 52 -

invention under the control of six copies of the GAL4 upstream activating sequences fused to the 35S minimal promoter. This is carried out using methods well known in the art. The trans-acting factor comprises the GAL4 DNA-binding domain (Keegan et al. (1986) *Science* 231: 699-704) fused to the transactivating domain of the herpes viral protein VP16 (Triezenberg et al. (1988) *Genes Devel.* 2: 718-729) fused to the hormone-binding domain of the rat glucocorticoid receptor (Picard et al. (1988) *Cell* 54: 1073-1080). The expression of the fusion protein is controlled by any promoter suitable for expression in plants known in the art or described here. This expression cassette is also comprised in the plant comprising a nucleic acid sequence of the invention fused to the 6xGAL4/minimal promoter. Thus, tissue- or organ-specificity of the fusion protein is achieved leading to inducible tissue- or organ-specificity of the insecticidal toxin.

g. Root Specific Expression:

Another pattern of gene expression is root expression. A suitable root promoter is described by de Framond (FEBS 290: 103-106 (1991)) and also in the published patent application EP 0 452 269, which is herein incorporated by reference. This promoter is transferred to a suitable vector such as pCGN1761ENX for the insertion of a selected gene and subsequent transfer of the entire promoter-gene-terminator cassette to a transformation vector of interest.

h. Wound-Inducible Promoters:

Wound-inducible promoters may also be suitable for gene expression. Numerous such promoters have been described (e.g. Xu et al. Plant Molec. Biol. 22: 573-588 (1993), Logemann et al. Plant Cell 1: 151-158 (1989), Rohrmeier & Lehle, Plant Molec. Biol. 22: 783-792 (1993), Firek et al. Plant Molec. Biol. 22: 129-142 (1993), Warner et al. Plant J. 3: 191-201 (1993)) and all are suitable for use with the instant invention. Logemann et al. describe the 5' upstream sequences of the dicotyledonous potato wunl gene. Xu et al. show that a wound-inducible promoter from the dicotyledon potato (pin2) is active in the monocotyledon rice. Further, Rohrmeier & Lehle describe the cloning of the maize Wipl cDNA which is wound induced and which can be used to isolate the cognate promoter using standard techniques. Similar, Firek et al. and Warner et al. have described a wound-induced gene from the monocotyledon Asparagus officinalis, which is expressed at local wound and pathogen invasion sites. Using cloning techniques well known in the art, these

promoters can be transferred to suitable vectors, fused to the genes pertaining to this invention, and used to express these genes at the sites of plant wounding.

i. Pith-Preferred Expression:

Patent Application WO 93/07278, which is herein incorporated by reference, describes the isolation of the maize *trpA* gene, which is preferentially expressed in pith cells. The gene sequence and promoter extending up to -1726 bp from the start of transcription are presented. Using standard molecular biological techniques, this promoter, or parts thereof, can be transferred to a vector such as pCGN1761 where it can replace the 35S promoter and be used to drive the expression of a foreign gene in a pith-preferred manner. In fact, fragments containing the pith-preferred promoter or parts thereof can be transferred to any vector and modified for utility in transgenic plants.

j. Leaf-Specific Expression:

A maize gene encoding phosphoenol carboxylase (PEPC) has been described by Hudspeth & Grula (Plant Molec Biol 12: 579-589 (1989)). Using standard molecular biological techniques the promoter for this gene can be used to drive the expression of any gene in a leaf-specific manner in transgenic plants.

k. Pollen-Specific Expression:

WO 93/07278 describes the isolation of the maize calcium-dependent protein kinase (CDPK) gene which is expressed in pollen cells. The gene sequence and promoter extend up to 1400 bp from the start of transcription. Using standard molecular biological techniques, this promoter or parts thereof, can be transferred to a vector such as pCGN1761 where it can replace the 35S promoter and be used to drive the expression of a nucleic acid sequence of the invention in a pollen-specific manner.

2. Transcriptional Terminators

A variety of transcriptional terminators are available for use in expression cassettes. These are responsible for the termination of transcription beyond the transgene and its correct polyadenylation. Appropriate transcriptional terminators are those that are known to function in plants and include the CaMV 35S terminator, the *tml* terminator, the nopaline synthase terminator and the pea *rbcS* E9 terminator. These can be used in both

monocotyledons and dicotyledons. In addition, a gene's native transcription terminator may be used.

3. Sequences for the Enhancement or Regulation of Expression

Numerous sequences have been found to enhance gene expression from within the transcriptional unit and these sequences can be used in conjunction with the genes of this invention to increase their expression in transgenic plants.

Various intron sequences have been shown to enhance expression, particularly in monocotyledonous cells. For example, the introns of the maize *Adhl* gene have been found to significantly enhance the expression of the wild-type gene under its cognate promoter when introduced into maize cells. Intron 1 was found to be particularly effective and enhanced expression in fusion constructs with the chloramphenicol acetyltransferase gene (Callis *et al.*, Genes Develop. 1: 1183-1200 (1987)). In the same experimental system, the intron from the maize *bronze1* gene had a similar effect in enhancing expression. Intron sequences have been routinely incorporated into plant transformation vectors, typically within the non-translated leader.

A number of non-translated leader sequences derived from viruses are also known to enhance expression, and these are particularly effective in dicotyledonous cells. Specifically, leader sequences from Tobacco Mosaic Virus (TMV, the "W-sequence"), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (e.g. Gallie et al. Nucl. Acids Res. 15: 8693-8711 (1987); Skuzeski et al. Plant Molec. Biol. 15: 65-79 (1990)).

4. Targeting of the Gene Product Within the Cell

Various mechanisms for targeting gene products are known to exist in plants and the sequences controlling the functioning of these mechanisms have been characterized in some detail. For example, the targeting of gene products to the chloroplast is controlled by a signal sequence found at the amino terminal end of various proteins which is cleaved during chloroplast import to yield the mature protein (*e.g.* Comai *et al.* J. Biol. Chem. <u>263</u>: 15104-15109 (1988)). These signal sequences can be fused to heterologous gene products to effect the import of heterologous products into the chloroplast (van den Broeck, et al. Nature <u>313</u>: 358-363 (1985)). DNA encoding for appropriate signal sequences can be isolated from the 5' end of the cDNAs encoding the RUBISCO protein, the CAB protein, the

EPSP synthase enzyme, the GS2 protein and many other proteins which are known to be chloroplast localized. *See also*, the section entitled "Expression With Chloroplast Targeting" in Example 37 of U.S. Patent No. 5,639,949.

Other gene products are localized to other organelles such as the mitochondrion and the peroxisome (e.g. Unger et al. Plant Molec. Biol. 13: 411-418 (1989)). The cDNAs encoding these products can also be manipulated to effect the targeting of heterologous gene products to these organelles. Examples of such sequences are the nuclear-encoded ATPases and specific aspartate amino transferase isoforms for mitochondria. Targeting cellular protein bodies has been described by Rogers et al. (Proc. Natl. Acad. Sci. USA 82: 6512-6516 (1985)).

In addition, sequences have been characterized which cause the targeting of gene products to other cell compartments. Amino terminal sequences are responsible for targeting to the ER, the apoplast, and extracellular secretion from aleurone cells (Koehler & Ho, Plant Cell 2: 769-783 (1990)). Additionally, amino terminal sequences in conjunction with carboxy terminal sequences are responsible for vacuolar targeting of gene products (Shinshi *et al.* Plant Molec. Biol. 14: 357-368 (1990)).

By the fusion of the appropriate targeting sequences described above to transgene sequences of interest it is possible to direct the transgene product to any organelle or cell compartment. For chloroplast targeting, for example, the chloroplast signal sequence from the RUBISCO gene, the CAB gene, the EPSP synthase gene, or the GS2 gene is fused in frame to the amino terminal ATG of the transgene. The signal sequence selected should include the known cleavage site, and the fusion constructed should take into account any amino acids after the cleavage site which are required for cleavage. In some cases this requirement may be fulfilled by the addition of a small number of amino acids between the cleavage site and the transgene ATG or, alternatively, replacement of some amino acids within the transgene sequence. Fusions constructed for chloroplast import can be tested for efficacy of chloroplast uptake by in vitro translation of in vitro transcribed constructions followed by in vitro chloroplast uptake using techniques described by Bartlett et al. In: Edelmann et al. (Eds.) Methods in Chloroplast Molecular Biology, Elsevier pp 1081-1091 (1982) and Wasmann et al. Mol. Gen. Genet. 205: 446-453 (1986). These construction techniques are well known in the art and are equally applicable to mitochondria and peroxisomes.

The above-described mechanisms for cellular targeting can be utilized not only in conjunction with their cognate promoters, but also in conjunction with heterologous promoters so as to effect a specific cell-targeting goal under the transcriptional regulation of a promoter that has an expression pattern different to that of the promoter from which the targeting signal derives.

Example 24: Construction of Plant Transformation Vectors

Numerous transformation vectors available for plant transformation are known to those of ordinary skill in the plant transformation arts, and the genes pertinent to this invention can be used in conjunction with any such vectors. The selection of vector will depend upon the preferred transformation technique and the target species for transformation. For certain target species, different antibiotic or herbicide selection markers may be preferred. Selection markers used routinely in transformation include the *nptll* gene, which confers resistance to kanamycin and related antibiotics (Messing & Vierra. Gene 19: 259-268 (1982); Bevan et al., Nature 304:184-187 (1983)), the *bar* gene, which confers resistance to the herbicide phosphinothricin (White et al., Nucl. Acids Res 18: 1062 (1990), Spencer et al. Theor. Appl. Genet 79: 625-631 (1990)), the *hph* gene, which confers resistance to the antibiotic hygromycin (Blochinger & Diggelmann, Mol Cell Biol 4: 2929-2931), and the *dhfr* gene, which confers resistance to methatrexate (Bourouis et al., EMBO J. 2(7): 1099-1104 (1983)), and the EPSPS gene, which confers resistance to glyphosate (U.S. Patent Nos. 4,940,935 and 5,188,642).

1. Vectors Suitable for Agrobacterium Transformation

Many vectors are available for transformation using *Agrobacterium tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan, Nucl. Acids Res. (1984)) and pXYZ. Below, the construction of two typical vectors suitable for *Agrobacterium* transformation is described.

a. pClB200 and pClB2001:

The binary vectors pclB200 and pClB2001 are used for the construction of recombinant vectors for use with *Agrobacterium* and are constructed in the following manner. pTJS75kan is created by *Narl* digestion of pTJS75 (Schmidhauser & Helinski, J.

PCT/EP99/01015

Bacteriol. 164: 446-455 (1985)) allowing excision of the tetracycline-resistance gene, followed by insertion of an Accl fragment from pUC4K carrying an NPTII (Messing & Vierra, Gene 19: 259-268 (1982): Bevan et al., Nature 304: 184-187 (1983): McBride et al., Plant Molecular Biology 14: 266-276 (1990)). Xhol linkers are ligated to the EcoRV fragment of PCIB7 which contains the left and right T-DNA borders, a plant selectable nos/nptll chimeric gene and the pUC polylinker (Rothstein et al., Gene 53: 153-161 (1987)), and the Xholdigested fragment are cloned into Sall-digested pTJS75kan to create pClB200 (see also EP 0 332 104, example 19). pCIB200 contains the following unique polylinker restriction sites: EcoRI, Sstl, Kpnl, Bglll, Xbal, and Sall. pClB2001 is a derivative of pClB200 created by the insertion into the polylinker of additional restriction sites. Unique restriction sites in the polylinker of pClB2001 are EcoRI, Sstl, KpnI, BgllI, XbaI, SalI, MluI, Bcll, AvrII, ApaI, HpaI, and Stul. pCIB2001, in addition to containing these unique restriction sites also has plant and bacterial kanamycin selection, left and right T-DNA borders for Agrobacterium-mediated transformation, the RK2-derived trfA function for mobilization between E. coli and other hosts, and the OriT and OriV functions also from RK2. The pCIB2001 polylinker is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

b. pCIB10 and Hygromycin Selection Derivatives thereof:

The binary vector pClB10 contains a gene encoding kanamycin resistance for selection in plants and T-DNA right and left border sequences and incorporates sequences from the wide host-range plasmid pRK252 allowing it to replicate in both *E. coli* and *Agrobacterium*. Its construction is described by Rothstein *et al.* (Gene <u>53</u>: 153-161 (1987)). Various derivatives of pClB10 are constructed which incorporate the gene for hygromycin B phosphotransferase described by Gritz *et al.* (Gene <u>25</u>: 179-188 (1983)). These derivatives enable selection of transgenic plant cells on hygromycin only (pClB743), or hygromycin and kanamycin (pClB715, pClB717).

2. Vectors Suitable for non-Agrobacterium Transformation

Transformation without the use of *Agrobacterium tumefaciens* circumvents the requirement for T-DNA sequences in the chosen transformation vector and consequently vectors lacking these sequences can be utilized in addition to vectors such as the ones described above which contain T-DNA sequences. Transformation techniques that do not rely on *Agrobacterium* include transformation via particle bombardment, protoplast uptake

(e.g. PEG and electroporation) and microinjection. The choice of vector depends largely on the preferred selection for the species being transformed. Below, the construction of typical vectors suitable for non-Agrobacterium transformation is described.

a. pCIB3064:

pCIB3064 is a pUC-derived vector suitable for direct gene transfer techniques in combination with selection by the herbicide basta (or phosphinothricin). The plasmid pCIB246 comprises the CaMV 35S promoter in operational fusion to the E. coli GUS gene and the CaMV 35S transcriptional terminator and is described in the PCT published application WO 93/07278. The 35S promoter of this vector contains two ATG sequences 5' of the start site. These sites are mutated using standard PCR techniques in such a way as to remove the ATGs and generate the restriction sites Sspl and Pvull. The new restriction sites are 96 and 37 bp away from the unique Sall site and 101 and 42 bp away from the actual start site. The resultant derivative of pCIB246 is designated pCIB3025. The GUS gene is then excised from pCIB3025 by digestion with Sall and Sacl, the termini rendered blunt and religated to generate plasmid pCIB3060. The plasmid pJIT82 is obtained from the John Innes Centre, Norwich and the a 400 bp Smal fragment containing the bar gene from Streptomyces viridochromogenes is excised and inserted into the Hpal site of pCIB3060 (Thompson et al. EMBO J 6: 2519-2523 (1987)). This generated pCIB3064, which comprises the bar gene under the control of the CaMV 35S promoter and terminator for herbicide selection, a gene for ampicillin resistance (for selection in E. coli) and a polylinker with the unique sites Sphl, Pstl, Hindlll, and BamHl. This vector is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

b. pSOG19 and pSOG35:

pSOG35 is a transformation vector that utilizes the *E. coli* gene dihydrofolate reductase (DFR) as a selectable marker conferring resistance to methotrexate. PCR is used to amplify the 35S promoter (-800 bp), intron 6 from the maize Adh1 gene (-550 bp) and 18 bp of the GUS untranslated leader sequence from pSOG10. A 250-bp fragment encoding the *E. coli* dihydrofolate reductase type II gene is also amplified by PCR and these two PCR fragments are assembled with a *SacI-PstI* fragment from pB1221 (Clontech) which comprises the pUC19 vector backbone and the nopaline synthase terminator. Assembly of these fragments generates pSOG19 which contains the 35S promoter in fusion

with the intron 6 sequence, the GUS leader, the DHFR gene and the nopaline synthase terminator. Replacement of the GUS leader in pSOG19 with the leader sequence from Maize Chlorotic Mottle Virus (MCMV) generates the vector pSOG35. pSOG19 and pSOG35 carry the pUC gene for ampicillin resistance and have *HindIII*, *SphI*, *PstI* and *EcoRI* sites available for the cloning of foreign substances.

Example 25: Transformation

Once a nucleic acid sequence of the invention has been cloned into an expression system, it is transformed into a plant cell. Methods for transformation and regeneration of plants are well known in the art. For example, Ti plasmid vectors have been utilized for the delivery of foreign DNA, as well as direct DNA uptake, liposomes, electroporation, microinjection, and microprojectiles. In addition, bacteria from the genus *Agrobacterium* can be utilized to transform plant cells. Below are descriptions of representative techniques for transforming both dicotyledonous and monocotyledonous plants.

1. Transformation of Dicotyledons

Transformation techniques for dicotyledons are well known in the art and include Agrobacterium-based techniques and techniques that do not require Agrobacterium. Non-Agrobacterium techniques involve the uptake of exogenous genetic material directly by protoplasts or cells. This can be accomplished by PEG or electroporation mediated uptake, particle bombardment-mediated delivery, or microinjection. Examples of these techniques are described by Paszkowski et al., EMBO J 3: 2717-2722 (1984), Potrykus et al., Mol. Gen. Genet. 199: 169-177 (1985), Reich et al., Biotechnology 4: 1001-1004 (1986), and Klein et al., Nature 327: 70-73 (1987). In each case the transformed cells are regenerated to whole plants using standard techniques known in the art.

Agrobacterium-mediated transformation is a preferred technique for transformation of dicotyledons because of its high efficiency of transformation and its broad utility with many different species. Agrobacterium transformation typically involves the transfer of the binary vector carrying the foreign DNA of interest (e.g. pCIB200 or pCIB2001) to an appropriate Agrobacterium strain which may depend of the complement of vir genes carried by the host Agrobacterium strain either on a co-resident Ti plasmid or chromosomally (e.g. strain CIB542 for pCIB200 and pCIB2001 (Uknes et al. Plant Cell 5: 159-169 (1993)). The

transfer of the recombinant binary vector to *Agrobacterium* is accomplished by a triparental mating procedure using *E. coli* carrying the recombinant binary vector, a helper *E. coli* strain which carries a plasmid such as pRK2013 and which is able to mobilize the recombinant binary vector to the target *Agrobacterium* strain. Alternatively, the recombinant binary vector can be transferred to *Agrobacterium* by DNA transformation (Höfgen & Willmitzer, Nucl. Acids Res. 16: 9877 (1988)).

Transformation of the target plant species by recombinant *Agrobacterium* usually involves co-cultivation of the *Agrobacterium* with explants from the plant and follows protocols well known in the art. Transformed tissue is regenerated on selectable medium carrying the antibiotic or herbicide resistance marker present between the binary plasmid T-DNA borders.

Another approach to transforming plant cells with a gene involves propelling inert or biologically active particles at plant tissues and cells. This technique is disclosed in U.S. Patent Nos. 4,945,050, 5,036,006, and 5,100,792 all to Sanford et al. Generally, this procedure involves propelling inert or biologically active particles at the cells under conditions effective to penetrate the outer surface of the cell and afford incorporation within the interior thereof. When inert particles are utilized, the vector can be introduced into the cell by coating the particles with the vector containing the desired gene. Alternatively, the target cell can be surrounded by the vector so that the vector is carried into the cell by the wake of the particle. Biologically active particles (e.g., dried yeast cells, dried bacterium or a bacteriophage, each containing DNA sought to be introduced) can also be propelled into plant cell tissue.

2. Transformation of Monocotyledons

Transformation of most monocotyledon species has now also become routine. Preferred techniques include direct gene transfer into protoplasts using PEG or electroporation techniques, and particle bombardment into callus tissue. Transformations can be undertaken with a single DNA species or multiple DNA species (*i.e.* cotransformation) and both these techniques are suitable for use with this invention. Cotransformation may have the advantage of avoiding complete vector construction and of generating transgenic plants with unlinked loci for the gene of interest and the selectable marker, enabling the removal of the selectable marker in subsequent generations, should this be regarded desirable. However, a disadvantage of the use of co-transformation is the

less than 100% frequency with which separate DNA species are integrated into the genome (Schocher et al. Biotechnology 4: 1093-1096 (1986)).

Patent Applications EP 0 292 435, EP 0 392 225, and WO 93/07278 describe techniques for the preparation of callus and protoplasts from an elite inbred line of maize, transformation of protoplasts using PEG or electroporation, and the regeneration of maize plants from transformed protoplasts. Gordon-Kamm *et al.* (Plant Cell <u>2</u>: 603-618 (1990)) and Fromm *et al.* (Biotechnology <u>8</u>: 833-839 (1990)) have published techniques for transformation of A188-derived maize line using particle bombardment. Furthermore, WO 93/07278 and Koziel *et al.* (Biotechnology <u>11</u>: 194-200 (1993)) describe techniques for the transformation of elite inbred lines of maize by particle bombardment. This technique utilizes immature maize embryos of 1.5-2.5 mm length excised from a maize ear 14-15 days after pollination and a PDS-1000He Biolistics device for bombardment.

Transformation of rice can also be undertaken by direct gene transfer techniques utilizing protoplasts or particle bombardment. Protoplast-mediated transformation has been described for *Japonica*-types and *Indica*-types (Zhang *et al.* Plant Cell Rep 7: 379-384 (1988); Shimamoto *et al.* Nature 338: 274-277 (1989); Datta *et al.* Biotechnology 8: 736-740 (1990)). Both types are also routinely transformable using particle bombardment (Christou *et al.* Biotechnology 9: 957-962 (1991)). Furthermore, WO 93/21335 describes techniques for the transformation of rice via electroporation.

Patent Application EP 0 332 581 describes techniques for the generation, transformation and regeneration of Pooideae protoplasts. These techniques allow the transformation of *Dactylis* and wheat. Furthermore, wheat transformation has been described by Vasil *et al.* (Biotechnology 10: 667-674 (1992)) using particle bombardment into cells of type C long-term regenerable callus, and also by Vasil *et al.* (Biotechnology 11: 1553-1558 (1993)) and Weeks *et al.* (Plant Physiol. 102: 1077-1084 (1993)) using particle bombardment of immature embryos and immature embryo-derived callus. A preferred technique for wheat transformation, however, involves the transformation of wheat by particle bombardment of immature embryos and includes either a high sucrose or a high maltose step prior to gene delivery. Prior to bombardment, any number of embryos (0.75-1 mm in length) are plated onto MS medium with 3% sucrose (Murashiga & Skoog, Physiologia Plantarum 15: 473-497 (1962)) and 3 mg/l 2,4-D for induction of somatic embryos, which is allowed to proceed in the dark. On the chosen day of bombardment, embryos are removed from the induction medium and placed onto the osmoticum (*i.e.*

induction medium with sucrose or maltose added at the desired concentration, typically 15%). The embryos are allowed to plasmolyze for 2-3 h and are then bombarded. Twenty embryos per target plate is typical, although not critical. An appropriate gene-carrying plasmid (such as pClB3064 or pSG35) is precipitated onto micrometer size gold particles using standard procedures. Each plate of embryos is shot with the DuPont Biolistics® helium device using a burst pressure of ~1000 psi using a standard 80 mesh screen. After bombardment, the embryos are placed back into the dark to recover for about 24 h (still on osmoticum). After 24 hrs, the embryos are removed from the osmoticum and placed back onto induction medium where they stay for about a month before regeneration. Approximately one month later the embryo explants with developing embryogenic callus are transferred to regeneration medium (MS + 1 mg/liter NAA, 5 mg/liter GA), further containing the appropriate selection agent (10 mg/l basta in the case of pClB3064 and 2 mg/l methotrexate in the case of pSOG35). After approximately one month, developed shoots are transferred to larger sterile containers known as "GA7s" which contain half-strength MS, 2% sucrose, and the same concentration of selection agent.

Tranformation of monocotyledons using *Agrobacterium* has also been described. *See,* WO 94/00977 and U.S. Patent No. 5,591,616, both of which are incorporated herein by reference.

E. Breeding and Seed Production

Example 26: Breeding

The plants obtained via tranformation with a nucleic acid sequence of the present invention can be any of a wide variety of plant species, including those of monocots and dicots; however, the plants used in the method of the invention are preferably selected from the list of agronomically important target crops set forth *supra*. The expression of a gene of the present invention in combination with other characteristics important for production and quality can be incorporated into plant lines through breeding. Breeding approaches and techniques are known in the art. See, for example, Welsh J. R., *Fundamentals of Plant Genetics and Breeding*, John Wiley & Sons, NY (1981); *Crop Breeding*, Wood D. R. (Ed.) American Society of Agronomy Madison, Wisconsin (1983); Mayo O., *The Theory of Plant Breeding*, Second Edition, Clarendon Press, Oxford (1987); Singh, D.P., *Breeding for*

- 63 -

Resistance to Diseases and Insect Pests, Springer-Verlag, NY (1986); and Wricke and Weber, Quantitative Genetics and Selection Plant Breeding, Walter de Gruyter and Co., Berlin (1986).

The genetic properties engineered into the transgenic seeds and plants described above are passed on by sexual reproduction or vegetative growth and can thus be maintained and propagated in progeny plants. Generally said maintenance and propagation make use of known agricultural methods developed to fit specific purposes such as tilling, sowing or harvesting. Specialized processes such as hydroponics or greenhouse technologies can also be applied. As the growing crop is vulnerable to attack and damages caused by insects or infections as well as to competition by weed plants, measures are undertaken to control weeds, plant diseases, insects, nematodes, and other adverse conditions to improve yield. These include mechanical measures such a tillage of the soil or removal of weeds and infected plants, as well as the application of agrochemicals such as herbicides, fungicides, gametocides, nematicides, growth regulants, ripening agents and insecticides.

Use of the advantageous genetic properties of the transgenic plants and seeds according to the invention can further be made in plant breeding, which aims at the development of plants with improved properties such as tolerance of pests, herbicides, or stress, improved nutritional value, increased yield, or improved structure causing less loss from lodging or shattering. The various breeding steps are characterized by well-defined human intervention such as selecting the lines to be crossed, directing pollination of the parental lines, or selecting appropriate progeny plants. Depending on the desired properties, different breeding measures are taken. The relevant techniques are well known in the art and include but are not limited to hybridization, inbreeding, backcross breeding, multiline breeding, variety blend, interspecific hybridization, aneuploid techniques, etc. Hybridization techniques also include the sterilization of plants to yield male or female sterile plants by mechanical, chemical, or biochemical means. Cross pollination of a male sterile plant with pollen of a different line assures that the genome of the male sterile but female fertile plant will uniformly obtain properties of both parental lines. Thus, the transgenic seeds and plants according to the invention can be used for the breeding of improved plant lines, that for example, increase the effectiveness of conventional methods such as herbicide or pestidice treatment or allow one to dispense with said methods due to their modified genetic properties. Alternatively new crops with improved stress tolerance can be obtained, which, due to their optimized genetic "equipment", yield harvested product of better quality than products that were not able to tolerate comparable adverse developmental conditions.

Example 27: Seed Production

In seed production, germination quality and uniformity of seeds are essential product characteristics, whereas germination quality and uniformity of seeds harvested and sold by the farmer is not important. As it is difficult to keep a crop free from other crop and weed seeds, to control seedborne diseases, and to produce seed with good germination, fairly extensive and well-defined seed production practices have been developed by seed producers, who are experienced in the art of growing, conditioning and marketing of pure seed. Thus, it is common practice for the farmer to buy certified seed meeting specific quality standards instead of using seed harvested from his own crop. Propagation material to be used as seeds is customarily treated with a protectant coating comprising herbicides, insecticides, fungicides, bactericides, nematicides, molluscicides, or mixtures thereof. Customarily used protectant coatings comprise compounds such as captan, carboxin, thiram (TMTD*), methalaxyl (Apron*), and pirimiphos-methyl (Actellic*). If desired, these compounds are formulated together with further carriers, surfactants or applicationpromoting adjuvants customarily employed in the art of formulation to provide protection against damage caused by bacterial, fungal or animal pests. The protectant coatings may be applied by impregnating propagation material with a liquid formulation or by coating with a combined wet or dry formulation. Other methods of application are also possible such as treatment directed at the buds or the fruit.

It is a further aspect of the present invention to provide new agricultural methods, such as the methods examplified above, which are characterized by the use of transgenic plants, transgenic plant material, or transgenic seed according to the present invention.

The seeds may be provided in a bag, container or vessel comprised of a suitable packaging material, the bag or container capable of being closed to contain seeds. The bag, container or vessel may be designed for either short term or long term storage, or both, of the seed. Examples of a suitable packaging material include paper, such as kraft paper, rigid or pliable plastic or other polymeric material, glass or metal. Desirably the bag, container, or vessel is comprised of a plurality of layers of packaging materials, of the same or differing type. In one embodiment the bag, container or vessel is provided so as to

exclude or limit water and moisture from contacting the seed. In one example, the bag, container or vessel is sealed, for example heat sealed, to prevent water or moisture from entering. In another embodiment water absorbent materials are placed between or adjacent to packaging material layers. In yet another embodiment the bag, container or vessel, or packaging material of which it is comprised is treated to limit, suppress or prevent disease, contamination or other adverse affects of storage or transport of the seed. An example of such treatment is sterilization, for example by chemical means or by exposure to radiation. Comprised by the present invention is a commercial bag comprising seed of a transgenic plant comprising a gene of the present invention that is expressed in said transformed plant at higher levels than in a wild type plant, together with a suitable carrier, together with label instructions for the use thereof for conferring broad spectrum disease resistance to plants.

BUDAPLET TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISHS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO

VIABILITY STATEMENT

Novartie AG Novartis Corporation 3054 Cornwallis Rd. Research Triangle Park, NC 27709

issued pursuant to Rule 10.2 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF THE PARTY TO WHOM THE VIRBILITY STATEMENT IS ISSUED

I. DEPOSITOR	II. IDENTIFICATION OF THE MICROORGANISM			
Name: Novartis AG Novartis Corporation Address: 3054 Cornwallis Rd. Research Triangle Park, NC 27709	Depositor's taxonomic designation and accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: Escherichia coli NRRL B-30077 Date of: October 28, 1998 X 2 Original Deposit 2 New Deposit 1 Repropagation of Original Deposit			
III. (a) VIABILITY STATEMENT				
Deposit was found: X Viable Nonviable on October 31, 1998 (Date) International Depositary Authority's preparation was found viable on December 8, 1998 (Date)				
III. (b) DEPOSITOR'S EQUIVALENCY DECLARA	ATION			
Depositor determined the International Dep				
2 Equivalent 2 Not equivalent to de signature of Depositor Note Note				
COMPLETONS UNDER WHICH THE VIABILITY	TEST WAS PERFORMED (Depositors/Depositary)			
IV. CONDITIONS UNDER WHICH THE VIABILITY TEST WAS PERFORMED (Depositors/Depositary). The dried culture was put into 2 mls LBampusughmi, and grown at 37°C overnight with shaking. Some of the liquid culture was streaked to an Lbampusughmi plate + grown at 37°C overnight.				
V. INTERNATIONAL DEPOSITARY AUTHORITY				
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority Address: 1815 N. University Street Peoria, Illinois 61604 U.S.A.	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s): 1			

Indicate the date of the original deposit or when a new deposit has been made.

Mark with a cross the applicable box.

In the cases referred to in Rule 10.2(a)(ii) and (iii), refer to the most recent viability test.

Fill in if the information has been requested.

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO
Novartis AG
Novartis Corporation
3054 Cornwallis Rd.
Research Triangle Park,
NC 27709

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT issued pursuant to Rule 7.1 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF DEPOSITOR

والمراجع والمراع والمراجع والم				
I. IDENTIFICATION OF THE MICROGRANISM				
Identification reference given by the DEPOSITOR:	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY:			
Ezcherichiz coli pNOV2400	NRL 8-30077			
II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION				
The microorganism identified under I. above was accompanied by:				
a scientific description				
x a proposed taxonomic designation				
(Mark with a cross where soplicable)				
III. RECEIPT AND ACCEPTANCE				
This International Depositary Authority accepts the microorganism identified under I. above, which was received by it on October 28, 1998(date of the original deposit)				
IV. RECEIPT OF REQUEST FOR CONVERSION				
The microorganism identified under I. above was received by this International Depositary Authority on (date of the original deposit) and a request to convert the original deposit to a deposit under the Eudapest Treaty was received by it on (date of receipt of request for conversion).				
V. INTERNATIONAL DEPOSITARY AUTHORITY				
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorited official(s):			
Address: 1815 N. University Street Peoria, Illinois 61604 U.S.A.	Date: 12 3-11			

Where Rule 6.4(d) applies, such date is the date on which the status f international depositary authority was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROGRAMISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL PORK

TO
Novertis AG
Novertis Corporation
3054 Cornwallis Rd.
Research Triangle Park,
NC 27709

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT issued pursuant to Rule 7.1 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF DEPOSITOR

1. IDENTIFICATION OF THE MICROORGANISM			
Identification reference given by the DEPOSITOR:	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY:		
Escherichia coli pNOV1001	HRRL 8-30078		
II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION			
The microorganism identified under I. above	ve was accompanied by:		
a scientific description			
x a proposed taxonomic designation			
(Mark with a cross where applicable)			
III. RECEIPT AND ACCEPTANCE			
This International Depositary Authority accepts the microorganism identified under I. above, which was received by it on October 28, 1998(date of the original deposit)			
IV. RECEIPT OF REQUEST FOR CONVERSION			
The microorganism identified under I. above was received by this International Depositary Authority on (date of the original deposit) and a request to convert the original deposit to a deposit under the Budupest Treaty was received by it on (date of receipt of request for conversion).			
V. INTERNATIONAL DEPOSITARY AUTHORITY			
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s):		
Address: 1815 N. University Street	Date: 12 3-7/		

^{&#}x27; where Rule 6.4(d) applies, such date is the date on which the status of international depositary authority was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION F THE DEPOSIT F MICROGRAMISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

VIABILITY STATEMENT

Novartis AG Novartis Corporation 3054 Cornwellis Rd. Research Triangle Park, NC 27709

issued pursuant to Rule 10.2 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF THE PARTY TO WHOM THE VIRBILITY STATEMENT IS ISSUED

I. DEPOSITOR	II. IDENTIFICATION OF THE MICROORGANISM			
Name: Novartis AG Novartis Corporation Address: 3054 Cornwallis Rd. Research Triangle Park, NC 27709	Depositor's taxonomic designation and accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: Escherichia coli NRRL B-30078 Date of: October 28, 1998 X 2 Original Deposit 2 New Deposit 3 Repropagation of Original Deposit			
III. (a) VIRBILITY STATEMENT				
Deposit was found: K Viable Nonviable on October 31, 1998 (Date) International Depositary Authority's preparation was found viable on December 8, 1996(Date)				
III. (b) DEPOSITOR'S EQUIVALENCY DECLAR	ATION			
Depositor determined the International Depositary Authority's preparation was Rquivalent Not equivalent to deposit on 1-6-99 (Date) Bignature of Depositor Hope Hant				
IV. CONDITIONS UNDER WHICH THE VINBILITY TEST WAS PERFORMED (Depositors/Depositary)				
IV. CONDITIONS UNDER WHICH THE VIABILITY TEST WAS PERFORMED (Depositors/Depositesty). The dried culture was put into and LBamp(longland) and grown at 37°C overnight with shaking. Some of the liquid culture was streaked to an LBamp plate and grown at 37°C overnight.				
V. INTERNATIONAL DEPOSITARY AUTHORITY				
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority Address: 1815 N. University Street	Signature(s) of person(s) having the power to represent the International Depositary Authority or of suthorized official(s):			
Peoris, Illinois 61604 U.S.A.	Dates			

Indicate the date of the original deposit or when a new deposit has been made.

*Mark with a cross the applicable box.

In the cases referred to in Rule 10.7(a)(ii) and (iii), refer to the most recent visbility test.

*Fill in if the information has been requested.

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURFOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO

Novartis Corp. c/o Novartis AG P. O. Box 12257

Research Triangle Park, NC 27709

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT issued pursuant to Rule 7.1 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF DEPOSITOR

I. IDENTIFICATION OF THE MICROORGANISM				
Identification reference given by the DEPOSITOR:	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY:			
Bacteria sp. pCIB 9359-7 NRRL B-21835				
II. SCIENTIFIC DESCRIPTION AND/OR PROPOS	ED TAXONOMIC DESIGNATION			
The microorganism identified under I. above was accompanied by:				
a scientific description				
a proposed taxonomic designation				
(Mark with a cross where applicable)				
III. RECEIPT AND ACCEPTANCE				
This International Depositary Authority accepts the microorganism identified under I. above, which was received by it on September 17, 1997 (date of the original deposit)'				
IV. RECEIPT OF REQUEST FOR CONVERSION				
The microorganism identified under I. above was received by this International Depositary Authority on (date of the original deposit) and a request to convert the original deposit to a deposit under the Budapest Treaty was received by it on (date of receipt of request for conversion).				
V. INTERNATIONAL DEPOSITARY AUTHORITY				
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority	Signature (s) of person (s) having the power to represent the International Depositary Authority or of authorized official (s):			
Address: 1815 N. University Street Peoria. Illinois 61604 U.S.A. Date:				

^{&#}x27;Where Rule 6.4(d) applies, such date is the date on which the status of international depositary authority was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO

VIABILITY STATEMENT

Novartis Corp. c/o Novartis AG P. O. Box 12357 Research Triangle Park, NC 27709

issued pursuant to Rule 10.2 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF THE PARTY TO WHOM THE VIABILITY STATEMENT IS ISSUED

THE VINDIBILL DISCLASSIVE TO LOCKED								
I. DEPOSITOR	II. IDENTIFICATION OF THE MICROORGANISM							
Name: Novartis Corp c/o Novartis AG Address: P. O. Box 12257 Research Triangle Park, NC 27709	Depositor's taxxnomic designation and accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: Bacteria sp. NRRL B-21835 Date of:September 17, 1997 2 Original Deposit 2 New Deposit: 2 Repropagation of Original Deposit							
III. (a) VIABILITY STATEMENT								
Deposit was found: Viable								
III. (b) DEFOSITOR'S EQUIVALENCY DECLARA	TION							
Depositor determined the International Dep	ositary Authority's preparation was							
: Equivalent : Not equivalent to de	posit on(Date)							
Signature of Dapositor								
IV. CONDITIONS UNDER WHICH THE VIABILITY	TEST WAS PERFORMED (Depositors/Depositary)							
V. INTERNATIONAL DEPOSITARY AUTHORITY								
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s):							
Address: 1815 N. University Street Peoria, Illinois 61604 U.S.A.	Pace: //- 17 · 77							

indicate the date of the original deposit or when a new deposit has been made.

Herk with a cross the applicable hea.

In the cases referred to in Rule 10.2(a)(11) and (111), refer to the most sacent viability test.

Fill in if the information has been requested.

What is claimed is:

- 1. An isolated nucleic acid molecule comprising:
 - (a) a nucleotide sequence substantially similar to a nucleotide sequence selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 15,171-18,035 of SEQ ID NO:11, and nucleotides 31,393-35,838 of SEQ ID NO:11;
 - (b) a nucleotide sequence comprising nucleotides 23,768-31,336 of SEQ ID NO:11; or
- (c) a nucleotide sequence isocoding with the nucleotide sequence of (a) or (b); wherein expression of said nucleic acid molecule results in at least one toxin that is active against insects.
- 2. An isolated nucleic acid molecule comprising a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of a nucleotide sequence selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 15,171-18,035 of SEQ ID NO:11, and nucleotides 31,393-35,838 of SEQ ID NO:11, wherein expression of said nucleic acid molecule results in at least one toxin that is active against insects.
- 3. An isolated nucleic acid molecule comprising a nucleotide sequence from *Photorhabdus luminescens* selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 66-1898 of SEQ ID NO:11, nucleotides 2416-9909 of SEQ ID NO:11, the complement of nucleotides 2817-3395 of SEQ ID NO:11, nucleotides 9966-14,633 of SEQ ID NO:11, nucleotides 14,699-15,007 of SEQ ID NO:11, nucleotides 15,171-18,035 of SEQ ID NO:11, the complement of nucleotides 17,072-17,398 of SEQ ID NO:11, the complement of nucleotides 19,385-20,116 of SEQ ID NO:11, the complement of nucleotides 20,217-20,963 of SEQ ID NO:11,

the complement of nucleotides 22,172-23,086 of SEQ ID NO:11, nucleotides 23,768-31,336 of SEQ ID NO:11, nucleotides 31,393-35,838 of SEQ ID NO:11, the complement of nucleotides 35,383-35,709 of SEQ ID NO:11, the complement of nucleotides 36,032-36,661 of SEQ ID NO:11, and the complement of nucleotides 36,654-37,781 of SEQ ID NO:11.

- 4. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence is substantially similar to nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1.
- 5. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence encodes an amino acid sequence selected from the group consisting of SEQ ID NOs:2-6.
- 6. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence comprises nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1.
- 7. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence is substantially similar to nucleotides 15,171-18,035 or 31,393-35,838 of SEQ ID NO:11.
- 8. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence encodes the amino acid sequence set forth in SEQ ID NOs:12-14.
- 9. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence comprises nucleotides 15,171-18,035; 23,768-31,336; or 31,393-35,838 of SEQ ID NO:11.
- 10. An isolated nucleic acid molecule according to claim 2, comprising a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides

2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1.

- 11. An isolated nucleic acid molecule according to claim 2, comprising a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of nucleotides 15,171-18,035 or 31,393-35,838 of SEQ ID NO:11.
- 12. A chimeric gene comprising a heterologous promoter sequence operatively linked to the nucleic acid molecule of claim 1 or claim 2.
- 13. A recombinant vector comprising the chimeric gene of claim 12.
- 14. A host cell comprising the chimeric gene of claim 12.
- 15. A host cell according to claim 14, which is a bacterial cell.
- 16. A host cell according to claim 14, which is a yeast cell.
- 17. A host cell according to claim 14, which is a plant cell.
- 18. A plant comprising the plant cell of claim 17.
- 19. A plant according to claim 18, which is maize.
- 20. A toxin produced by the expression of a DNA molecule according to claim 1 or claim 2.
- 21. A toxin according to claim 20, wherein said toxin has activity against Lepidopteran insects.
- 22. A toxin according to claim 21, wherein said toxin has activity against *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm).

- 23. A toxin according to claim 20, wherein said toxin has activity against Lepidopteran and Coleopteran insects.
- 24. A toxin according to claim 23, wherein said toxin has insecticidal activity against Plutella xylostella (Diamondback Moth), Ostrinia nubilalis (European Corn Borer), and Manduca sexta (Tobacco Hornworm), Diabrotica virgifera virgifera (Western Corn Rootworm), Diabrotica undecimpunctata howardi (Southern Corn Rootworm), and Leptinotarsa decimlineata (Colorado Potato Beetle).
- 25. A toxin according to claim 20, wherein said toxin comprises an amino acid sequence selected from the group consisting of: SEQ ID NOs:2-6.
- 26. A toxin according to claim 20, wherein said toxin comprises an amino acid sequence selected from the group consisting of: SEQ ID NOs:12-14.
- 27. A composition comprising an insecticidally effective amount of a toxin according to claim 20.
- 28. A method of producing a toxin that is active against insects, comprising:
 - (a) obtaining the host cell of claim 14; and
 - (b) expressing the nucleic acid molecule in said cell, which results in at least one toxin that is active against insects.
- 29. A method of producing an insect-resistant plant, comprising introducing a nucleic acid molecule according to claim 1 into said plant, wherein said nucleic acid molecule is expressible in said plant in an effective amount to control insects.
- 30. A method of controlling insects comprising delivering to the insects an effective amount of a toxin according to claim 44.
- 31. The method of claim 29 or claim 30, wherein the insects are Lepidopteran insects.

- 32. The method of claim 31, wherein the insects are selected from the group consisting of: *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm).
- 33. The method of claim 29 or claim 30, wherein the insects are Lepidopteran and Coleopteran insects.
- 34. The method of claim 33, wherein the insects are selected from the group consisting of: *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle).
- 35. The method of claim 30, wherein the toxin is delivered to the insects orally.
- 36. A method for mutagenizing a nucleic acid molecule according to claim 1, wherein the nucleic acid molecule has been cleaved into population of double-stranded random fragments of a desired size, comprising:
 - (a) adding to the population of double-stranded random fragments one or more single- or double-stranded oligonucleotides, wherein said oligonucleotides each comprise an area of identity and an area of heterology to a doublestranded template polynucleotide;
 - (b) denaturing the resultant mixture of double-stranded random fragments and oligonucleotides into single-stranded fragments;
 - (c) incubating the resultant population of single-stranded fragments with a polymerase under conditions which result in the annealing of said single-stranded fragments at said areas of identity to form pairs of annealed fragments, said areas of identity being sufficient for one member of a pair to prime replication of the other, thereby forming a mutagenized double-stranded polynucleotide; and

(d) repeating the second and third steps for at least two further cycles, wherein the resultant mixture in the second step of a further cycle includes the mutagenized double-stranded polynucleotide from the third step of the previous cycle, and wherein the further cycle forms a further mutagenized double-stranded polynucleotide.

```
SEQUENCE LISTING
<110> Novartis AG
<120> Novel Toxins And Uses Thereof
<130> PI/5-30421/A/CGC 1963
<140>
<141>
<160> 22
<170> PatentIn Ver. 2.0
<210> 1
<211> 9717
<212> DNA
<213> Photorhabdus luminescens
<220>
<221> CDS
<222> (412)..(1665)
<223> orf1 ~46.4 kDa
<220>
<221> CDS
<222> (1686)..(2447)
<223> orf2 ~28.1kDa
<220>
<221> CDS
<222> (2758)..(3318)
<223> orf3 ~20.7 kDa
<220>
<221> CDS
<222> (3342)..(4118)
<223> orf4 ~28.7 kDa
<220>
<221> CDS
<222> (4515)..(9269)
<223> orf5 ~176 kDa
<400> 1
gaattcatat gctatgaaat aaacagttgg cgcaataatt aaagctatta tttttatttt 60
gtttttatac aatgatatgc tttattaaac agaataatga gttaatgata aataaatcct 120
cgggatttat catgatatta tggccgaatg tgatgtgaac aattatttta taattagatt 180
aataatataa tggtattaaa ataacaatat atttattcat gggtatttat catcggtttt 240
attacatggg gaataatcta taaattagtt ttacataatt cacaaatagc gattccatta 300
 accaggaata ttaaaaatac ttatttatga ttatggtgat atatcttcat tagcctactt 360
 ttataactag aaaaattgac attttcaatc catgtataaa tggtaaccaa t atg cag 417
                                                          Met Gln
 aga get caa ega gtt gtt att aca ggt atg ggt gee gta aca eeg att
 Arg Ala Gln Arg Val Val Ile Thr Gly Met Gly Ala Val Thr Pro Ile
```

ggt Gly	gaa Glu 20	gat Asp	gtt Val	gaa Glu	tca Ser	tgt Cys 25	tgg Trp	caa Gln	agt Ser	att Ile	att Ile 30	gaa Glu	aaa Lys	caa Gln	cat His	513
cga Arg 35	ttt Phe	cac His	aga Arg	att Ile	gaa Glu 40	ttt Phe	cct Pro	gac Asp	tca Ser	ttc Phe 45	att Ile	aat Asn	tcg Ser	cgt. Arg	ttc Phe 50	561
ttt Phe	tct Ser	ttc Phe	ctt Leu	gca Ala 55	cca Pro	aac Asn	cca Pro	tcc Ser	cgc Arg 60	tat Tyr	cag Gln	tta Leu	tta Leu	cca Pro 65	aaa Lys	609
aag Lys	ttg Leu	act Thr	cat His 70	aca Thr	ctt Leu	tct Ser	gac Asp	tgc Cys 75	gga Gly	aaa Lys	gca Ala	gca Ala	ttg Leu 80	aag Lys	gcg Ala	657
act Thr	tat Tyr	caa Gln 85	gct Ala	ttt Phe	acc Thr	caa Gln	gca Ala 90	ttc Phe	ggc Gly	gtg Val	aat Asn	ata Ile 95	tca Ser	cct Pro	gtt Val	705
gaa Glu	tat Tyr 100	tac Tyr	gat Asp	aaa Lys	tac Tyr	gaa Glu 105	tgt Cys	Gly ggc	gta Val	att Ile	ctt Leu 110	ggc Gly	agt Ser	ggt Gly	tgg Trp	753
gga Gly 115	gct Ala	att Ile	gat Asp	aat Asn	gcc Ala 120	gga Gly	gat Asp	cat His	gct Ala	tgc Cys 125	caa Gln	tat Tyr	aag Lys	caa Gln	gca Ala 130	801
aaa Lys	tta Leu	gct Ala	cat His	cct Pro 135	atg Met	agt Ser	aat Asn	ctt Leu	att Ile 140	acc Thr	atg Met	cca Pro	agc Ser	tcc Ser 145	atg Met	849
				Ser				gga Gly 155								897
			Ala										Asp		ttt Phe	945
		Ile										Ala			gct Ala	993
gaa Glu 195	Ser	cto Leu	acg Thr	. Arg	gaa Glu 200	Cys	aat Asn	att Ile	tgg Trp	agt Ser 205	Ile	gat Asp	gta Val	ctg Leu	aat Asn 210	1041
					: Gln					Leu					ttt Phe	1089
agc Ser	ctt Leu	gat Asp	cgc Arg 230	, Ser	gga Gly	ttt Phe	gta Val	Leu 235	ı Ala	gaa Glu	a gga u Gly	a gcc / Ala	gcg Ala 240	Va]	gtt Val	1137
			ı Asr					ı Ile					Thr		tta Leu	1185
gcg	gaa Glu 260	ı Ile	aaa Lys	a ggt s Gly	tac Tyn	gcc Ala 265	Gli	a tat n Tyn	tca Ser	gat Asp	gco Ala 270	a Val	aat LAsr	tta Leu	a acc ı Thr	1233

egg Arg 275	cca Pro	aca Thr	gaa Glu	gat Asp	att Ile 280	gaa Glu	cct Pro	aaa Lys	ata Ile	tta Leu 285	gcg Ala	ata Ile	act Thr	aaa Lys	gcc Ala 290	1281
att Ile	gag Glu	cag Gln	gca Ala	cag Gln 295	att Ile	tcg Ser	ccg Pro	aaa Lys	gat Asp 300	att Ile	gac Asp	tac Tyr	att Ile	aat Asn 305	gct Ala	1329
cat His	ggt Gly	act Thr	tct Ser 310	aca Thr	ccg Pro	tta Leu	aat Asn	gat Asp 315	ctt Leu	tat Tyr	gaa Glu	act Thr	cag Gln 320	gca Ala	att Ile	1377
aaa Lys	gca Ala	gca Ala 325	ctg Leu	ggc Gly	caa Gln	tat Tyr	gct Ala 330	tat Tyr	cag Gln	gta Val	cct Pro	ata Ile 335	tca Ser	agc Ser	aca Thr	1425
aaa Lys	tct Ser 340	tat Tyr	acc Thr	ggc Gly	cac His	ctt Leu 345	att Ile	gct Ala	gcc Ala	gcc Ala	ggt Gly 350	agt Ser	ttt Phe	gaa Glu	acg Thr	1473
att Ile 355	gta Val	tgt Cys	gtg Val	aaa Lys	gca Ala 360	tta Leu	gct Ala	gaa Glu	aat Asn	tgc Cys 365	ttg Leu	cca Pro	gca Ala	aca Thr	ttg Leu 370	1521
aat Asn	tta Leu	cac His	cgg Arg	gcc Ala 375	gat Asp	cca Pro	gat. Asp	tgc Cys	gat Asp 380	ctc Leu	aat Asn	tat Tyr	ttg Leu	cct Pro 385	aat Asn	1569
caa Gln	cat His	tgc Cys	tac Tyr 390	acc Thr	gct Ala	caa Gln	cca Pro	gag Glu 395	gtg Val	aca Thr	ctc Leu	aat Asn	att Ile 400	agc Ser	gca Ala	1617
ggt Gly	ttc Phe	ggc Gly 405	ggg ggg	cat His	aac Asn	gct Ala	gcg Ala 410	ttg Leu	gtt Val	atc Ile	gct Ala	aag Lys 415	gta Val	agg Arg	taa	1665
GIY	Phe	G1y 405	Gly	His	Asn caat	Ala :g ga	Ala 410 aa ga lu As	Leu at at	Val :t ga	Ile aa ca	Ala at to	Lys 415 g to p Se	Val eg aa	Arg	taa cc tct ne Ser	
ctga	Phe	GIY 405 gtt g	Gly gattt aac	His ttgo	Asn ca at Me	Ala g ga et Gl 42 cat	Ala 410 aa ga lu As 20 tat	Leu at at sp Il	Val it ga le Gl	Ile aa ca lu Hi	Ala at to is Tr 42	Lys 415 gg to p Se 25	Val eg aa er As	Arg	tc tct ne Ser	
ctga ggg Gly 430	Phe atato gat	Gly 405 gtt g ttt Phe	Gly gattt aac Asn caa	ttgc ccc Pro	Asn ca at Me atc Ile 435	Ala g ga et Gl 42 cat His	Ala 410 aa ga lu As 20 tat Tyr	Leu at at sp I] tcg Ser	Val t ga le Gl gcg Ala	Ile aa ca lu Hi aaa Lys 440	Ala at to is Tr 42 agc Ser	Lys 415 gg to p Se 25 gag Glu	Val eg aa er As tet Ser	Arg at tt sn Pl ttg Leu	cgc Arg 445	1718
ctga ggg Gly 430 aat Asn	Phe atato gat Asp ata	gtt g ttt Phe cag Gln	gatti aac Asn caa Gln	ccc Pro cac His 450	Asn atc Ile 435 ccg Pro	Ala g ga et Gl 42 cat His gtg Val	Ala 410 aa ga lu As 20 tat Tyr cag Gln	Leu at at sp Il tcg Ser gga Gly	Val t ga le Gl gcg Ala atg Met 455	Ile aa ca lu Hi aaa Lys 440 ttg Leu	Ala at to is Tr 42 agc Ser agt Ser	Lys 415 gg to p Se 25 gag Glu ttg Leu	Val eg aæ tet As tet Ser etc Leu	Arg at the sn Ph ttg Leu tat Tyr 460	cgc tet cgc Arg 445 gta Val	1718 1766
ggg Gly 430 aat Asn cgg Arg	gat Asp ata Ile	ttt Phe cag Gln cag Gln	aac Asn caa Gln ttt Phe 465 gcc	ccc Pro cac His 450 tct Ser	Asn atc atc lle 435 ccg Pro caa Gln	Ala cg ga et Gl 42 cat His gtg Val tta Leu cgc	Ala 410 410 tat Tyr cag Gln act Thr	Leu tcg Ser gga Gly tcc Ser 470 tat	Val tt ga le Gl gcg Ala atg Met 455 gct Ala	Ile aa ca lu Hi aaa Lys 440 ttg Leu ttt Phe	Ala At tous Tr 42 agc Ser agt Ser aca Thr	Lys 415 415 gg tc pp Se 55 gag Glu ttg Leu acg Thr	Val gg aa tct Ser ctc Leu gga 475	Arg at the trong plant of the trong plant tat Tyr 460 ata Ile	cgc Arg 445 gta Val ttg Leu	1718 1766 1814
ggg Gly 430 aat Asn cgg Arg	gat Asp ata Ile caa Gln att	ttt check the cag Gln cag Gln gat Asp 480 agg	aac Asn caa Gln ttt Phe 465 gcc Ala att	ccc Pro cac His 450 tct Ser tct Ser	Asn atc Ile 435 ccg Pro caa Gln ttc Phe act	Ala cg ga cat His gtg Val tta Leu cgc Arg	Ala 410 Ala	Leu tcg Il tcg Ser gga Gly tcc Ser 470 tat Tyr	Val tt ga le Gl gcg Ala atg Met 455 gct Ala gtt Val	Ile Ile Ile Ile Ily Ile Itt Phe Itt Tyr	Ala at too is Tr 42 agc Ser agt Ser aca Thr acc Thr	Lys 415 gg tc pp Se 25 gag Glu ttg Leu acg Thr gca Ala 490	Val cg ag cr As tct Ser ctc Leu gga Gly 475 tta Leu	Arg at the sn Ph ttg Leu tat Tyr 460 ata Ile ccc Pro	cgc tct cgc Arg 445 gta Val ttg Leu cat His	1718 1766 1814 1862
ggg Gly 430 aat Asn cgg Arg aac Asn caa Gln agt	gat Asp ata Ile caa Gln att Ile	Gly 405 ttt Phe cag Gln cag Gln gat Asp 480 agg Arg	aac Asn caa Gln ttt Phe 465 gcc Ala att lle aac	ccc Pro cac His 450 tct Ser tct Ser aat Asn	Asn ca at Me atc Ile 435 ccg Pro caa Gln ttc Phe act Thr	Ala Ig ga Ig ga Ig ga Ala Cat Gl Ala Cat His Gtg Val tta Leu Cgc Arg aaa Lys 5000 ttc	Ala 410 Ala 420 tat Tyr cag Gln act Thr cag Ass	Leu tcg Ser gga Gly tcc Ser 470 tat Tyr aaa Lys	Val tt ga le Gl gcg Ala atg Met 455 gct Ala gtt Val acg Thr	Ile Ile Ile Ile Ilys 440 Ittg Leu Itt Phe Itt Tyr Itt Phe agc	Ala at to ser 42 agc Ser agt Ser aca Thr acc Thr aaa Lys 505 ota	Lys 415 gg tc pp Se 25 gag Glu ttg Leu acg Thr gca Ala 490 tta Leu	Val cg ag er Ag tct Ser ctc Leu gga Gly 475 tta Leu ggaa Glu aat	Arg at the sin Pl ttg Leu tat Tyr 460 ata Ile ccc Pro aat	cc tct ne Ser cgc Arg 445 gta Val ttg Leu cat His	1718 1766 1814 1862 1910

Glu	Ser	Ile	Glu	Asp 530	Trp	Ile	Val	Gln	Asp 535	Asn	Cys	Gln	Lys	Leu 540	Thr	
ata Ile	aca Thr	Gly ggg	gag Glu 545	gaa Glu	gtt Val	tgt Cys	gaa Glu	aag Lys 550	tat Tyr	gct Ala	gtc Val	ttt Phe	aga Arg 555	tac Tyr	tat Tyr	2102
ttc Phe	cca Pro	agt Ser 560	gtc Val	act Thr	tct Ser	att Ile	gga Gly 565	tgg Trp	ttc Phe	ctg Leu	gat Asp	gcg Ala 570	ctt Leu	gct Ala	ttt Phe	2150
cat His	ctt Leu 575	att Ile	att Ile	aat Asn	tcg Ser	aca Thr 580	gga Gly	ttt Phe	ctt Leu	aat Asn	ttt Phe 585	gag Glu	cac His	tac Tyr	cat His	2198
ttt Phe 590	aac Asn	caa Gln	tta Leu	cag Gln	gat Asp 595	tat Tyr	ctg Leu	agt Ser	caa Gln	tct Ser 600	ttt Phe	act Thr	ttg Leu	cat His	act Thr 605	2246
ggg Gly	caa Gln	gcg Ala	att Ile	aaa Lys 610	atc Ile	agg Arg	aag Lys	gag Glu	att Ile 615	gtt Val	aat Asn	agt Ser	aca Thr	gta Val 620	tta Leu	2294
tta Leu	tct Ser	tca Ser	ccg Pro 625	gat Asp	atc Ile	tgt Cys	gtt Val	gaa Glu 630	tta Leu	aat Asn	cct Pro	cct Pro	tta Leu 635	ttg Leu	att Ile	2342
aag Lys	aat Asn	ggc Gly 640	gat Asp	aaa Lys	gat Asp	tat Tyr	att Ile 645	cgt Arg	att Ile	ttc Phe	tat Tyr	tat Tyr 650	cga Arg	tgt Cys	tta Leu	2390
tat Tyr	gat Asp 655	aaa Lys	aaa Lys	cct Pro	att Ile	ttt Phe 660	gta Val	tca Ser	aag Lys	act Thr	tca Ser 665	att Ile	atc Ile	tct Ser	aag Lys	2438
	aaa Lys	taa	aagg	gaaaq	gcg a	aatq	gccaa	ac ad	caaag	gtgat	ati	ttca	actg			2487
aaat	caaag	gaa t	tagaa	atati	ta at	gato	gaagg	g ata	ataga	aga	tgaa	agaaa	ata a	acaco	cagagt	2547
cct	cttt	igt 1	tcg	ettga	aa tt	tgat	agto	c ttg	gacta	atgt	ggaa	atco	caa q	gttt	tgtgt	2607
tgga	aagcg	gta 1	tggta	attg	cg ct	taaa	agcc	g aad	ettt	ttc	aaat	catt	cct a	attto	caacat	2667
taaa	atgag	gct (cacto	gacta	at t	caaaa	atcaa	a aat	tgta	atc	tgaa	attt	ta d	cttaa	attatg	2727
ttt	ttca	acc a	attaa	acati	ca aç	gaggt	tata	a ato Met	gaad Ast	gtt n Val 679	l Le	a gaz ı Glu	a caa ı Glr	a ggt n Gly	aag Lys 680	2781
gtt Val	gct Ala	gct Ala	tta Leu	tat Tyr 685	tca Ser	gcc Ala	tat Tyr	tcg Ser	gaa Glu 690	aca Thr	gaa Glu	ggt Gly	tct Ser	tcg Ser 695	tgg Trp	2829
gtg Val	gga Gly	aac Asn	ttg Leu 700	tgc Cys	tgt Cys	ttt Phe	tca Ser	agt Ser 705	gat Asp	cgg Arg	gag Glu	cat His	ttg Leu 710	cct Pro	att Ile	2877
atc Ile	gtg Val	aat Asn 715	ggg Gly	cgt Arg	cgt Arg	ttc Phe	ttg Leu 720	att Ile	gaa Glu	ttt Phe	gtt Val	att Ile 725	cca Pro	gat Asp	cat His	2925
tta Leu	ctt Leu	gat Asp	aaa Lys	acg Thr	gtt Val	aaa Lys	ccc Pro	aga Arg	gta Val	ttc Phe	gat Asp	ttg Leu	gat Asp	atc Ile	aat Asn	2973

730	735			740			
aaa caa ttt tta d Lys Gln Phe Leu 1 745							3021
tta ggt gaa gga a Leu Gly Glu Gly a		Asp Arg					3069
ttc gag tta aat (Phe Glu Leu Asn (780							3117
gct ctt ggt aaa Ala Leu Gly Lys 795	tat gtt gct Tyr Val Ala	att aat Ile Asn 800	cct tca Pro Ser	act acg Thr Thr 805	caa ttt Gln Phe	atc Ile	3165
Phe Phe Ala Gln 6 810							3213
aca gtt gaa gac Thr Val Glu Asp 825							3261
gaa ttg caa ggg Glu Leu Gln Gly	ccg tat tat Pro Tyr Tyr 845	Gly Phe	gaa ctt Glu Leu 850	gat att Asp Ile	ctt tct Leu Ser 855	att Ile	3309
aca gct taa ttca Thr Ala	caatat tatg	gagagt gt	Met G		ys Ile T	hr Thr	3362
			860		8	65	
ttt acc att gag Phe Thr Ile Glu 870	aaa act gat Lys Thr Asp	gac aat Asp Asn 875	ttt tat	gct aat Ala Asn	ggg cgt	cat	3410
Phe Thr Ile Glu	Lys Thr Asp aaa atc tct	Asp Asn 875 gta ctt	ttt tat Phe Tyr aaa caa	Ala Asn gaa tat	ggg cgt Gly Arg 880 agg aat Arg Asn	cat His	3410 3458
Phe Thr Ile Glu 870 caa tgt atg gta Gln Cys Met Val	Lys Thr Asp aaa atc tct Lys Ile Ser tta gca ctt	Asp Asn 875 gta ctt Val Leu 890 agt gag	ttt tat Phe Tyr aaa caa Lys Gln gct gaa	gaa tat Glu Tyr 895	ggg cgt Gly Arg 880 agg aat Arg Asn	cat His ggt Gly	
Phe Thr Ile Glu 870 caa tgt atg gta Gln Cys Met Val 885 gat tgg ata aaa Asp Trp Ile Lys	Lys Thr Asp aaa atc tct Lys Ile Ser tta gca ctt Leu Ala Leu 905 agt gat agc	Asp Asn 875 gta ctt Val Leu 890 agt gag Ser Glu ctc ata	ttt tat Phe Tyr aaa caa Lys Gln gct gaa Ala Glu tat gac	gaa tat Glu Tyr 895 aaa aga Lys Arg 910	ggg cgt Gly Arg 880 agg aat Arg Asn tcg att Ser Ile	cat His ggt Gly cag Gln	3458
Phe Thr Ile Glu 870 caa tgt atg gta Gln Cys Met Val 885 gat tgg ata aaa Asp Trp Ile Lys 900 gtg gcg gca tta Val Ala Ala Leu 915 tca ggt tgg aca Ser Gly Trp Thr	Lys Thr Asp aaa atc tct Lys Ile Ser tta gca ctt Leu Ala Leu 905 agt gat agc Ser Asp Ser 920 acg aca gat	Asp Asn 875 gta ctt Val Leu 890 agt gag Ser Glu ctc ata Leu Ile gca aga	ttt tat Phe Tyr aaa caa Lys Gln gct gaa Ala Glu tat gac Tyr Asp 925 aat aaa	gaa tat Glu Tyr 895 aaa aga Lys Arg 910 caa tta Gln Leu	ggg cgt Gly Arg 880 agg aat Arg Asn tcg att Ser Ile aaa atg Lys Met	cat His ggt Gly cag Gln cct Pro 930 tta Leu	3458 3506
Phe Thr Ile Glu 870 caa tgt atg gta Gln Cys Met Val 885 gat tgg ata aaa Asp Trp Ile Lys 900 gtg gcg gca tta Val Ala Ala Leu 915 tca ggt tgg aca Ser Gly Trp Thr	Lys Thr Asp aaa atc tct Lys Ile Ser tta gca ctt Leu Ala Leu 905 agt gat agc Ser Asp Ser 920 acg aca gat Thr Thr Asp 935 tat cat gct	Asp Asn 875 gta ctt Val Leu 890 agt gag Ser Glu ctc ata Leu Ile gca aga Ala Arg	ttt tat Phe Tyr aaa caa Lys Gln gct gaa Ala Glu tat gac Tyr Asp 925 aat aaa Asn Lys 940 ttt att	gaa tat Glu Tyr 895 aaa aga Lys Arg 910 caa tta Gln Leu ttt gat Phe Asp	ggg cgt Gly Arg 880 agg aat Arg Asn tcg att Ser Ile aaa atg Lys Met ctt ggg Leu Gly 945 cag gta	cat His ggt Gly cag Gln cct Pro 930 tta Leu	3458 3506 3554
Phe Thr Ile Glu 870 caa tgt atg gta Gln Cys Met Val 885 gat tgg ata aaa Asp Trp Ile Lys 900 gtg gcg gca tta Val Ala Ala Leu 915 tca ggt tgg aca Ser Gly Trp Thr tta aat ggt gtt Leu Asn Gly Val	Lys Thr Asp aaa atc tct Lys Ile Ser tta gca ctt Leu Ala Leu 905 agt gat agc Ser Asp Ser 920 acg aca gat Thr Thr Asp 935 tat cat gct Tyr His Ala gat tgc tgc	Asp Asn 875 gta ctt Val Leu 890 agt gag Ser Glu ctc ata Leu Ile gca aga Ala Arg gat gct Asp Ala 955 aca aat	ttt tat Phe Tyr aaa caa Lys Gln gct gaa Ala Glu tat gac Tyr Asp 925 aat aaa Asn Lys 940 ttt att Phe Ile	gaa tat Glu Tyr 895 aaa aga Lys Arg 910 caa tta Gln Leu ttt gat Phe Asp gac gaa Asp Glu tat cag	ggg cgt Gly Arg 880 agg aat Arg Asn tcg att Ser Ile aaa atg Lys Met ctt ggg Leu Gly 945 cag gta Gln Val 960 aac agt Asn Ser	cat His ggt Gly cag Gln cct Pro 930 tta Leu aca Thr	3458 3506 3554 3602

agc Ser 995	aca Thr	gaa Glu	tac Tyr	Leu	atg Met .000	gca Ala	aaa Lys	atg Met	Thr	Phe 1005	gaa Glu	gat Asp	acg Thr	gat Asp 1	ggg Gly .010	3794
aaa Lys	cgc Arg	aca Thr	Leu	aca Thr 1015	acg Thr	aat Asn	atg Met	Ser	gtt Val 1020	ggt Gly	gat Asp	gaa Glu	Val	ttt Phe 1025	gac Asp	3842
agc Ser	aag Lys	Val	tta Leu 1030	tta Leu	aaa Lys	gcc Ala	Ile	gct Ala .035	cct Pro	tat Tyr	gca Ala	Ile	aat Asn .040	aca Thr	aat Asn	3890
caa Gln	Leu	cat His 1045	gaa Glu	aac Asn	atc Ile	Asn	aca Thr 1050	ttg Leu	ttt Phe	gat Asp	Lys	aca Thr 1055	gaa Glu	gag Glu	ccg Pro	3938
Thr	aaa Lys .060	tcc Ser	gat Asp	act Thr	His	cat His 1065	caa Gln	ata Ile	att Ile	Asn	ctt Leu 1070	tat Tyr	cgc Arg	tgg Trp	aca Thr	3986
ttg Leu 1075	Pro	tat Tyr	cat His	Leu	agg Arg 1080	att Ile	ctt Leu	gaa Glu	Gly	aat Asn 1085	gac Asp	agt Ser	act Thr	gtt Val	aat Asn 1090	4034
			Val					Pro					Phe	ctg Leu 1105		4082
aga Arg	gga Gly	Arg	gta Val 1110	ttt Phe	aaa Lys	cga Arg	gga Gly	act Thr 1115	cat His	atg Met	tga	atg	cacg	tga		4128
taat	gtg	agt	ggag	gatg	tg t	tatg	gact	a tg	ctta	tacc	gta	acta	ttc	cgga	cacgca	4188
gcti	tgct	gct	gaag	tgct	tc a	tgtg	acag	g gt	gttc	gtgg	acg	agtg	gtt	atta	tgatgg	4248
ata	cat	gat	gtca	caat	ca t	tgat	aact	a cg	gttg	tcag	cat	aaat	tta	gaat	ttcttc	4308
ggti	taat	att	ggac	gtgc	gc t	aagc	atag	c ga	gaat	aagt	tga	tttt	cct	tagt	aaaaaa	4368
cct	ttgt	tta	tgct	ggta	aa c	gcat	gtgc	g tt	tgcc	agca	att	aata	tat	tcca	ttattg	4428
aaa	tagg	aat	atag	ccat	at c	tgta	atta	t ac	ataa	acga	att	ttta	ctc	gaat	ataatt	4488
tta	attg	atc	aaac	agga	aa t	ttaa			s Al					r Se	c aat r Asn	4541
			Phe					Asn					Pro	aga Arg		4589
ĞÎy		туг					Asn					Arg		aat Asn	aat Asn	4637
	Gly					Thr					Phe				aca Thr 1175	4685
					⁄ Ph∈					Arg					aca Thr	4733

tta gat ata aaa aca ctt aca ttt agc cga gca aat ggg gag caa ttt Leu Asp Ile Lys Thr Leu Thr Phe Ser Arg Ala Asn Gly Glu Gln Phe 1195 1200 1205	4781
aaa tgt aag cca ttg ccg cct aat aat aat gat ctt agt ttt aaa gat Lys Cys Lys Pro Leu Pro Pro Asn Asn Asn Asp Leu Ser Phe Lys Asp 1210 1215 1220	4829
aaa aaa cta aaa gat ttg cgc gta tat aag ctc gat agc aat act ttt Lys Lys Leu Lys Asp Leu Arg Val Tyr Lys Leu Asp Ser Asn Thr Phe 1225 1230 1235	4877
tat gtt tat aac aaa aac ggc att ata gag ata ctt aaa cga att ggg Tyr Val Tyr Asn Lys Asn Gly Ile Ile Glu Ile Leu Lys Arg Ile Gly 1240 1245 1250 1250	•
tcg agt gat att gca aaa aca gtt gca ctt gaa ttt cct gat ggt gaa Ser Ser Asp Ile Ala Lys Thr Val Ala Leu Glu Phe Pro Asp Gly Glu 1260 1265 1270	1 4973 1
gca ttt gat tta att tat aat tca aga ttt gca ttg tcc gaa ata aaa Ala Phe Asp Leu Ile Tyr Asn Ser Arg Phe Ala Leu Ser Glu Ile Lys 1275 1280 1285	
tac cgt gtg aca ggt aaa act tat ctt aaa ctc aat tac tct gga aat Tyr Arg Val Thr Gly Lys Thr Tyr Leu Lys Leu Asn Tyr Ser Gly Asr 1290 1295 1300	
aac tgt aca tca gtg gaa tac cct gat gat aat aat att tct gcg aaa Asn Cys Thr Ser Val Glu Tyr Pro Asp Asn Asn Ile Ser Ala Lys 1305 1310 1315	a 5117 s
ata gca ttc gat tat cgt aac gat tac ctt att acg gtg act gta cci Ile Ala Phe Asp Tyr Arg Asn Asp Tyr Leu Ile Thr Val Thr Val Pro 1320 1325 1330 1339	5
tac gat gct tct ggt cct att gat tct gcc cga ttt aag atg acc ta Tyr Asp Ala Ser Gly Pro Ile Asp Ser Ala Arg Phe Lys Met Thr Ty 1340 1345 1350	t 5213 r
cag aca tta aaa ggc gta ttt cca gtt atc agc acc ttc cgt aca cc Gln Thr Leu Lys Gly Val Phe Pro Val Ile Ser Thr Phe Arg Thr Pr 1355 1360 1365	
acc ggt tat gtt gag ctg gtg agt tat aaa gag aat ggg cat aaa gt Thr Gly Tyr Val Glu Leu Val Ser Tyr Lys Glu Asn Gly His Lys Va 1370 1375 1380	g 5309 1
acg gac acg gaa tat att cct tat gcg gct gca ctc act att caa cc Thr Asp Thr Glu Tyr Ile Pro Tyr Ala Ala Ala Leu Thr Ile Gln Pr 1385 1390 1395	
ggc aat gga caa cct gcg gtc agc aaa tcc tat gaa tat agt tca gt Gly Asn Gly Gln Pro Ala Val Ser Lys Ser Tyr Glu Tyr Ser Ser Va 1400 1405 1410 141	1
cat aac ttc ttg ggc tat tct tct ggc cgg aca agc ttt gat tcc ag His Asn Phe Leu Gly Tyr Ser Ser Gly Arg Thr Ser Phe Asp Ser Se 1420 1425 1430	
caa gat aat ttg tat ttg gtc aca ggg aaa tac act tat tca tcc at Gln Asp Asn Leu Tyr Leu Val Thr Gly Lys Tyr Thr Tyr Ser Ser Il 1435 1440 1445	
gaa cgg gtt tta gat ggt caa agt gtg gtt tca gta ata gaa cga gt	a 5549

Glu Arg Val Leu Asp Gly Gln Ser Val Val Ser Val Ile Glu Arg Val 1450 1455 1460	
ttt aat aaa ttc cat tta atg acc aaa gaa gca aaa aca caa gat aat Phe Asn Lys Phe His Leu Met Thr Lys Glu Ala Lys Thr Gln Asp Asn 1465 1470 1475	5597
aag aga att aca aca gaa att act tac aat gag gat cta tca aaa agt Lys Arg Ile Thr Thr Glu Ile Thr Tyr Asn Glu Asp Leu Ser Lys Ser 1480 1485 1490 1495	5645
ttc tca gag caa cca gaa aat tta caa caa cct tct cgc gtg tta acc Phe Ser Glu Gln Pro Glu Asn Leu Gln Gln Pro Ser Arg Val Leu Thr 1500 1505 1510	5693
cgt tat acg gat ata caa aca aat act tca cga gaa gag act gtc aat Arg Tyr Thr Asp Ile Gln Thr Asn Thr Ser Arg Glu Glu Thr Val Asn 1515 1520 1525	5741
att aaa agt gat gat tgg gga aat act cta ctt att act gag acc agt Ile Lys Ser Asp Asp Trp Gly Asn Thr Leu Leu Ile Thr Glu Thr Ser 1530 1535 1540	5789
ggg ata cag aaa gaa tac gtt tat tat ccg gtc aat ggc gaa ggt aat Gly Ile Gln Lys Glu Tyr Val Tyr Tyr Pro Val Asn Gly Glu Gly Asn 1545 1550 1555	5837
agt tgc cct gcc gat ccc ttg ggt ttt tct cgg ttc tta aaa tca gtt Ser Cys Pro Ala Asp Pro Leu Gly Phe Ser Arg Phe Leu Lys Ser Val 1560 1565 1570 1575	5885
acg caa aaa gga tcg cct gat gct gct caa agt gtc gca aat aaa gtg Thr Gln Lys Gly Ser Pro Asp Ala Ala Gln Ser Val Ala Asn Lys Val 1580 1585 1590	5933
att cat tat aca tat caa aaa ttt cct act ttt acc ggc gct tat gtt Ile His Tyr Thr Tyr Gln Lys Phe Pro Thr Phe Thr Gly Ala Tyr Val 1595 1600 1605	5981
aag gaa tat gtc agt aaa gtc tca gag acg ata gac aat aaa ata gcg Lys Glu Tyr Val Ser Lys Val Ser Glu Thr Ile Asp Asn Lys Ile Ala 1610 1615 1620	6029
aga acc ttt agc tat gtt aac tca ccg acg agt aaa tct cat ggt tcg Arg Thr Phe Ser Tyr Val Asn Ser Pro Thr Ser Lys Ser His Gly Ser 1625 1630 1635	6077
tta gca aaa ata acg tca gtg atg aat aac cag caa acg gtc acc aca Leu Ala Lys Ile Thr Ser Val Met Asn Asn Gln Gln Thr Val Thr 1640 1645 1650 1655	6125
ttt aaa tat gaa tat toa gaa agt gag atg acc aca aat got acg gtg Phe Lys Tyr Glu Tyr Ser Glu Ser Glu Met Thr Thr Asn Ala Thr Val 1660 1665 1670	6173
acc ggt ttt gat ggc gca cat atg gaa tcg aaa aat gtg acg tct att Thr Gly Phe Asp Gly Ala His Met Glu Ser Lys Asn Val Thr Ser Ile 1675 1680 1685	6221
tat acc cat cgg caa ctt cgt aaa gtt gat gta aac cac gtg att acc Tyr Thr His Arg Gln Leu Arg Lys Val Asp Val Asn His Val Ile Thr 1690 1695 1700	6269
gat cag tot tat gat out tig ggt ogo att aca ggg caa att att gat Asp Gln Ser Tyr Asp Leu Leu Gly Arg Ile Thr Gly Gln Ile Ile Asp	6317

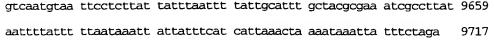
17	705				1	710				1	.715					
ccc g Pro G 1720	gc Sly	acg Thr	gca Ala	Arg	gaa Glu .725	att Ile	aaa Lys	cgt Arg	Asn	tac Tyr .730	gtt Val	tat Tyr	caa Gln	Tyr	ccc Pro 1735	6365
ggc g Gly G			Glu					Pro					Val			6413
caa g Gln G		Val					His					Gly				6461
tcg a Ser I	[le					Asp					Gly					6509
tat c Tyr C					Arg					Arg						6557
ggg (Gly (1800				Lys					Asp					Leu		6605
gcc a Ala A			Leu					Thr					Thr			6653
tat a Tyr I		Tyr					Asn					Glu				6701
ggt (Gly)	Arg		Glu			Ile					Thr					6749
caa (Gln (_			Leu		Met			Ile		Gln				6797
gag Glu 1880	Gln			Ser		Lys			Tyr		Asp			Ile		6845
agc Ser			Thr		Arg			Gly		Gly			Val		Glu	6893
				Gly			Thr		Ile					Phe	gat Asp	6941
	Ile		Lys					Āsp					Glu		gct Ala	6989
Tyr		Ser					ı Ğlu					Leu			aat Asn	7037
	Thr					Leu					Leu				ata Ile 1975	7085

agt gat acg gtg ggt ggt cgc aaa acg gaa tat tta tat ggg cct caa Ser Asp Thr Val Gly Gly Arg Lys Thr Glu Tyr Leu Tyr Gly Pro Gln 1980 1985 1990	7133
ggt gac aaa ccg att cag tca att act cct tcg cat aat aag caa aat Gly Asp Lys Pro Ile Gln Ser Ile Thr Pro Ser His Asn Lys Gln Asn 1995 2000 2005	7181
atg gat tac ctc tac tat ctt ggt agt gtg atg tcc aaa ttt acc acg Met Asp Tyr Leu Tyr Tyr Leu Gly Ser Val Met Ser Lys Phe Thr Thr 2010 2015 2020	7229
ggg aca gac caa caa aac ttt cgt tat cat tcg aaa acg gga aca tta Gly Thr Asp Gln Gln Asn Phe Arg Tyr His Ser Lys Thr Gly Thr Leu 2025 2030 2035	7277
tta tct gcg tca gaa ggc gta tct cag act aat tac agt tat ttc cca Leu Ser Ala Ser Glu Gly Val Ser Gln Thr Asn Tyr Ser Tyr Phe Pro 2040 2045 2050 2055	7325
tcg ggt gta tta cag cga gaa tca ttt tta cgg gat aat aaa ccg att Ser Gly Val Leu Gln Arg Glu Ser Phe Leu Arg Asp Asm Lys Pro Ile 2060 2065 2070	7373
tca tcg ggc gag tac ctt tat acg atg tcc ggt ttg att caa cgt cat Ser Ser Gly Glu Tyr Leu Tyr Thr Met Ser Gly Leu Ile Gln Arg His 2075 2080 2085	7421
aaa gat agt ttt ggt cat aat cat gtt tat agt tac gat gct cag gga Lys Asp Ser Phe Gly His Asn His Val Tyr Ser Tyr Asp Ala Gln Gly 2090 2095 2100	7469
aga ttg gtc aaa aca gaa cag gat gca caa tac gct aca ttt gaa tat Arg Leu Val Lys Thr Glu Gln Asp Ala Gln Tyr Ala Thr Phe Glu Tyr 2105 2110 2115	7517
gac aat gtt ggg cga ttg ata aca acg acg acc aaa gac acg acg tca Asp Asn Val Gly Arg Leu Ile Thr Thr Thr Thr Lys Asp Thr Thr Ser 2120 2125 2130 2135	7565
tta tcc caa tta gtg aca aaa atc gaa tat gat gct ttt gat cga gaa Leu Ser Gln Leu Val Thr Lys Ile Glu Tyr Asp Ala Phe Asp Arg Glu 2140 2145 2150	7613
ata aaa cgc tcg cta att agt gac ttc tca ata caa gtt att acc tta Ile Lys Arg Ser Leu Ile Ser Asp Phe Ser Ile Gln Val Ile Thr Leu 2155 2160 2165	7661
agc tat acg aag aat aat caa atc agt caa cgt atc acc tcc atc gat Ser Tyr Thr Lys Asn Asn Gln Ile Ser Gln Arg Ile Thr Ser Ile Asp 2170 2175 2180	7709
ggg gtg gtt atg aaa aat gaa cgt tat caa tat gat aat aat caa cgc Gly Val Val Met Lys Asn Glu Arg Tyr Gln Tyr Asp Asn Asn Gln Arg 2185 2190 2195	7757
tta agc caa tac caa tgt gag gga gaa caa tct ccg att gat cat acg Leu Ser Gln Tyr Gln Cys Glu Gly Glu Gln Ser Pro Ile Asp His Thr 2200 2205 2210 2215	7805
ggt cgt gta tta aat cag cag att tac cat tat gac caa tgg gga aat Gly Arg Val Leu Asn Gln Gln Ile Tyr His Tyr Asp Gln Trp Gly Asn 2220 2230	7853

att aag cgg ctc gat aat aca tat cga gat ggt aag gaa acg gtg gat Ile Lys Arg Leu Asp Asn Thr Tyr Arg Asp Gly Lys Glu Thr Val Asp 2235 2240 2245	7901
tat cat ttc agt caa gcc gat cca act caa ctt att cgt att acc agc Tyr His Phe Ser Gln Ala Asp Pro Thr Gln Leu Ile Arg Ile Thr Ser 2250 2255 2260	7949
gac aaa cag cag ata gag tta agt tat gat gct aat ggc aac cta aca Asp Lys Gln Gln Ile Glu Leu Ser Tyr Asp Ala Asn Gly Asn Leu Thr 2265 2270 2275	7997
cgt gac gaa aaa ggg caa acg ctc att tac gat cag aat aat cgc ttg Arg Asp Glu Lys Gly Gln Thr Leu Ile Tyr Asp Gln Asn Asn Arg Leu 2280 2285 2290 2295	8045
gta cag gtc aaa gac cgg ttg ggc aat ctg gtg tgc agc tac cag tat Val Gln Val Lys Asp Arg Leu Gly Asn Leu Val Cys Ser Tyr Gln Tyr 2300 2305 2310	8093
gat gca ttg aac aaa tta acc gca cag gtt ttg gcg aat ggt acc gtt Asp Ala Leu Asm Lys Leu Thr Ala Gln Val Leu Ala Asm Gly Thr Val 2315 2320 2325	8141
aat cga cag cat tat gct tcc ggt aaa gtg acg aat att caa ttg ggt Asn Arg Gln His Tyr Ala Ser Gly Lys Val Thr Asn Ile Gln Leu Gly 2330 2335 2340	8189
gat gaa gcg att act tgg ttg agc agt gat aag caa cga att gga cat Asp Glu Ala Ile Thr Trp Leu Ser Ser Asp Lys Gln Arg Ile Gly His 2345 2350 2355	8237
caa agc gcc aag aat ggt caa tca gtc tac tat caa tat ggt att gac Gln Ser Ala Lys Asn Gly Gln Ser Val Tyr Tyr Gln Tyr Gly Ile Asp 2360 2365 2370 2375	8285
cat aac agt acg gtt atc gcc agt cag aac gaa aac gag ttg atg gct His Asn Ser Thr Val Ile Ala Ser Gln Asn Glu Asn Glu Leu Met Ala 2380 2385 2390	8333
tta tcc tat aca cct tat ggc ttt agg agt tta att tcc tca tta ccg Leu Ser Tyr Thr Pro Tyr Gly Phe Arg Ser Leu Ile Ser Ser Leu Pro 2395 2400 2405	8381
ggt ttg aat ggc gca cag gtt gat cca gta aca ggc tgg tac ttc tta Gly Leu Asn Gly Ala Gln Val Asp Pro Val Thr Gly Trp Tyr Phe Leu 2410 2415 2420	8429
ggt aac gga tat cgt gtt ttc aac ccg gtt ctc atg agg ttt cac agc Gly Asn Gly Tyr Arg Val Phe Asn Pro Val Leu Met Arg Phe His Ser 2425 2430 2435	8477
ccc gat agt tgg agt cct ttt ggt cgg gga ggg att aac cct tat accPro Asp Ser Trp Ser Pro Phe Gly Arg Gly Gly Ile Asn Pro Tyr Thr2440244524502455	8525
tat tgc caa ggc gat ccc ata aac cgg att gat ctg aac ggt cat ctt Tyr Cys Gln Gly Asp Pro Ile Asn Arg Ile Asp Leu Asn Gly His Leu 2460 2465 2470	8573
agt gcc ggc ggg ata tta ggc att gtg cta ggg gca att ggc atc att Ser Ala Gly Gly Ile Leu Gly Ile Val Leu Gly Ala Ile Gly Ile Ile 2475 2480 2485	8621
gtc ggg att gta tca ctg gga gcc gga gcg gcg att agc gcg ggt ctc	8669



Val Gly Ile Val Ser Leu Gly Ala Gly Ala Ala Ile Ser Ala Gly Leu 2490 2495 2500	
att gct gcg ggg ggc gct ttg ggg gcg att gct tct acc agc gcg ctt Ile Ala Ala Gly Gly Ala Leu Gly Ala Ile Ala Ser Thr Ser Ala Leu 2505 2510 2515	۱7
gca gtt act gcg act gtc att gga ttg gct gcc gat tcg ata ggg att Ala Val Thr Ala Thr Val Ile Gly Leu Ala Ala Asp Ser Ile Gly Ile 2520 2535 2530 2535	55
gcg tca gca gca tta tcg gaa aaa gat ccg aaa aca tct ggg ata tta 881 Ala Ser Ala Ala Leu Ser Glu Lys Asp Pro Lys Thr Ser Gly Ile Leu 2540 2545 2550	13
aat tgg att agt gcg gga ttg ggg gtt tta agc ttt ggt atc agc gca 886 Asn Trp Ile Ser Ala Gly Leu Gly Val Leu Ser Phe Gly Ile Ser Ala 2555 2560 2565	51
ata acc ttt acc tct tcg ctg gta aaa tcg gca cgg agt ggt tct cag Ile Thr Phe Thr Ser Ser Leu Val Lys Ser Ala Arg Ser Gly Ser Gln 2570 2575 2580	39
gca gtc agc gcg ggt gtt atc ggg tca gtg cct ctt gaa ttt ggt gaa 899 Ala Val Ser Ala Gly Val Ile Gly Ser Val Pro Leu Glu Phe Gly Glu 2585 2590 2595	57
gtt gct agc cgt tcc agc aga cga tgg gat att gcg tta tct tcg ata Val Ala Ser Arg Ser Ser Arg Arg Trp Asp Ile Ala Leu Ser Ser Ile 2600 2605 2610 2615	05
teg ttg ggc gca aat gcg gcg tet ete tet acg ggg ata gcg gcg gcg Ser Leu Gly Ala Asn Ala Ala Ser Leu Ser Thr Gly Ile Ala Ala Ala 2620 2625 2630	53
gcg gtt gca gac agt aat gcg aat gca gct aat att ctg gga tgg gta 91 Ala Val Ala Asp Ser Asn Ala Asn Ala Asn Ile Leu Gly Trp Val 2635 2640 2645	01
tcc ttt ggt ttt ggt gca gta tcg aca acc tca gga ata att gag ctt 91 Ser Phe Gly Phe Gly Ala Val Ser Thr Thr Ser Gly Ile Ile Glu Leu 2650 2655 2660	.49
acg cgt aca gct tat gca gtg aat cat cag act tgg gaa ctg agt tca 91 Thr Arg Thr Ala Tyr Ala Val Asn His Gln Thr Trp Glu Leu Ser Ser 2665 2670 2675	.97
tca gca ggt act tcg gag gaa gtg aag cct ata cgt tgt ctc gtt tca 92 Ser Ala Gly Thr Ser Glu Glu Val Lys Pro Ile Arg Cys Leu Val Ser 2680 2685 2690 2695	245
cac cgc tgg aat cag aag cag tga atgttaaccc tcctcgggca gttgagttaa 92 His Arg Trp Asn Gln Lys Gln 2700	299
tcaaacgttt cgaaatagta ccgggaacta tttagccaat cgtccattga aacccgtaat 93	159
gtgttgcgac gtcgtttgac aatataaaga ttctgcgaac cgattggtta agtctcacga 94	119
aaaataacta ttaggcgaca tttgcgtcgc cttttttaag gaactttatc aggttacatt 94	179
tataagaagc tattttgttt tcgacggatg ttggtttctc tgagataaaa aatagaggga 95	539
aatgatgtca agggtgataa tggttaattg taaaatatgt gatattattc gcatttatat 95	599



<210> 2 <211> 417 <212> PRT <213> Photorhabdus luminescens <400> 2 Met Gln Arg Ala Gln Arg Val Val Ile Thr Gly Met Gly Ala Val Thr Pro Ile Gly Glu Asp Val Glu Ser Cys Trp Gln Ser Ile Ile Glu Lys 20 25 30Gln His Arg Phe His Arg Ile Glu Phe Pro Asp Ser Phe Ile Asn Ser Arg Phe Phe Ser Phe Leu Ala Pro Asn Pro Ser Arg Tyr Gln Leu Leu Pro Lys Lys Leu Thr His Thr Leu Ser Asp Cys Gly Lys Ala Ala Leu 65 70 75 80 Lys Ala Thr Tyr Gln Ala Phe Thr Gln Ala Phe Gly Val Asn Ile Ser Pro Val Glu Tyr Tyr Asp Lys Tyr Glu Cys Gly Val Ile Leu Gly Ser Gly Trp Gly Ala Ile Asp Asn Ala Gly Asp His Ala Cys Gln Tyr Lys Gln Ala Lys Leu Ala His Pro Met Ser Asn Leu Ile Thr Met Pro Ser Ser Met Thr Ala Ala Cys Ser Ile Met Tyr Gly Leu Arg Gly Tyr Gln 155 Asn Thr Val Met Ala Ala Cys Ala Thr Gly Thr Met Ala Ile Gly Asp Ala Phe Glu Ile Ile Arg Ser Gly Arg Ala Lys Cys Met Ile Ala Gly Ala Ala Glu Ser Leu Thr Arg Glu Cys Asn Ile Trp Ser Ile Asp Val Leu Asn Ala Leu Ser Lys Glu Gln Ala Asp Pro Asn Leu Ala Cys Cys Pro Phe Ser Leu Asp Arg Ser Gly Phe Val Leu Ala Glu Gly Ala Ala Val Val Cys Leu Glu Asn Tyr Asp Ser Ala Ile Ala Arg Gly Ala Thr Ile Leu Ala Glu Ile Lys Gly Tyr Ala Gln Tyr Ser Asp Ala Val Asn

Leu Thr Arg Pro Thr Glu Asp Ile Glu Pro Lys Ile Leu Ala Ile Thr

Lys Ala Ile Glu Gln Ala Gln Ile Ser Pro Lys Asp Ile Asp Tyr Ile 295 Asn Ala His Gly Thr Ser Thr Pro Leu Asn Asp Leu Tyr Glu Thr Gln Ala Ile Lys Ala Ala Leu Gly Gln Tyr Ala Tyr Gln Val Pro Ile Ser Ser Thr Lys Ser Tyr Thr Gly His Leu Ile Ala Ala Ala Gly Ser Phe Glu Thr Ile Val Cys Val Lys Ala Leu Ala Glu Asn Cys Leu Pro Ala Thr Leu Asn Leu His Arg Ala Asp Pro Asp Cys Asp Leu Asn Tyr Leu Pro Asn Gln His Cys Tyr Thr Ala Gln Pro Glu Val Thr Leu Asn Ile 395 Ser Ala Gly Phe Gly Gly His Asn Ala Ala Leu Val Ile Ala Lys Val Arg <210> 3 <211> 253 <212> PRT <213> Photorhabdus luminescens <400> 3 Met Glu Asp Ile Glu His Trp Ser Asn Phe Ser Gly Asp Phe Asn Pro Ile His Tyr Ser Ala Lys Ser Glu Ser Leu Arg Asn Ile Gln Gln His Pro Val Gln Gly Met Leu Ser Leu Leu Tyr Val Arg Gln Gln Phe Ser Gln Leu Thr Ser Ala Phe Thr Thr Gly Ile Leu Asn Ile Asp Ala Ser Phe Arg Gln Tyr Val Tyr Thr Ala Leu Pro His Gln Leu Arg Ile Asn Thr Lys Asn Lys Thr Phe Lys Leu Glu Asn Pro Ser Lys Glu Asn Thr Leu Phe Gly Asn Thr Ser Val Glu Asn Thr Met Glu Ser Ile Glu Asp 105 Trp Ile Val Gln Asp Asn Cys Gln Lys Leu Thr Ile Thr Gly Glu Glu 120 Val Cys Glu Lys Tyr Ala Val Phe Arg Tyr Tyr Phe Pro Ser Val Thr Ser Ile Gly Trp Phe Leu Asp Ala Leu Ala Phe His Leu Ile Ile Asn

Ser Thr Gly Phe Leu Asn Phe Glu His Tyr His Phe Asn Gln Leu Gln

170

Asp Tyr Leu Ser Gln Ser Phe Thr Leu His Thr Gly Gln Ala Ile Lys 180 185 190

Ile Arg Lys Glu Ile Val Asn Ser Thr Val Leu Leu Ser Ser Pro Asp 195 200 205

Ile Cys Val Glu Leu Asn Pro Pro Leu Leu Ile Lys Asn Gly Asp Lys 210 215 220

Asp Tyr Ile Arg Ile Phe Tyr Tyr Arg Cys Leu Tyr Asp Lys Lys Pro 225 230 235 240

Ile Phe Val Ser Lys Thr Ser Ile Ile Ser Lys Met Lys 245 250

<210> 4

<211> 186

<212> PRT

<213> Photorhabdus luminescens

<400> 4

Met Asn Val Leu Glu Gln Gly Lys Val Ala Ala Leu Tyr Ser Ala Tyr

1 5 10 15

Ser Glu Thr Glu Gly Ser Ser Trp Val Gly Asn Leu Cys Cys Phe Ser 20 25 30

Ser Asp Arg Glu His Leu Pro Ile Ile Val Asn Gly Arg Arg Phe Leu 35 40 45

Ile Glu Phe Val Ile Pro Asp His Leu Leu Asp Lys Thr Val Lys Pro 50 60

Arg Val Phe Asp Leu Asp Ile Asn Lys Gln Phe Leu Leu Arg Arg Asp 65 70 75 80

His Arg Glu Ile Asn Ile Tyr Leu Leu Gly Glu Gly Asn Phe Met Asp 85 90 95

Arg Thr Thr Asp Lys Asn Leu Phe Glu Leu Asn Glu Asp Gly Ser 100 105 110

Leu Phe Ile Lys Thr Leu Arg His Ala Leu Gly Lys Tyr Val Ala Ile 115 120 125

Asn Pro Ser Thr Thr Gln Phe Ile Phe Phe Ala Gln Gly Lys Tyr Ser 130 140

Glu Phe Ile Met Asn Ala Leu Lys Thr Val Glu Asp Glu Leu Ser Lys 145 150 155 160

Arg Tyr Arg Val Arg Ile Ile Pro Glu Leu Gln Gly Pro Tyr Tyr Gly 165 170 175

Phe Glu Leu Asp Ile Leu Ser Ile Thr Ala 180 185

<210> 5

<211> 258

<212> PRT

<213> Photorhabdus luminescens

- 16 -

<400> 5 Met Glu Lys Lys Ile Thr Thr Phe Thr Ile Glu Lys Thr Asp Asp Asn Phe Tyr Ala Asn Gly Arg His Gln Cys Met Val Lys Ile Ser 15 20 25 30 Val Leu Lys Gln Glu Tyr Arg Asn Gly Asp Trp Ile Lys Leu Ala Leu Ser Glu Ala Glu Lys Arg Ser Ile Gln Val Ala Ala Leu Ser Asp Ser Leu Ile Tyr Asp Gln Leu Lys Met Pro Ser Gly Trp Thr Thr Asp Ala Arg Asn Lys Phe Asp Leu Gly Leu Leu Asn Gly Val Tyr His Ala Asp Ala Phe Ile Asp Glu Gln Val Thr Asp Arg Ala Gly Asp Cys Cys Thr Asn Glu Asn Tyr Gln Asn Ser Val Lys Ser Val Pro Glu Ile Ile Tyr Arg Tyr Val Ser Ser Asn Arg Thr Ser Thr Glu Tyr Leu Met Ala Lys Met Thr Phe Glu Asp Thr Asp Gly Lys Arg Thr Leu Thr Thr Asn Met Ser Val Gly Asp Glu Val Phe Asp Ser Lys Val Leu Leu Lys Ala Ile Ala Pro Tyr Ala Ile Asn Thr Asn Gln Leu His Glu Asn Ile Asn Thr Leu Phe Asp Lys Thr Glu Glu Pro Thr Lys Ser Asp Thr His His 200 Gln Ile Ile Asn Leu Tyr Arg Trp Thr Leu Pro Tyr His Leu Arg Ile Leu Glu Gly Asn Asp Ser Thr Val Asn Arg Ile Tyr Val Leu Gly Lys 230 Glu Pro Ser Asn Asp Arg Phe Leu Thr Arg Gly Arg Val Phe Lys Arg Gly Thr His Met <210> 6 <211> 1584 <212> PRT <213> Photorhabdus luminescens <400> 6 Met Lys Ala Thr Asp Ile Tyr Ser Asn Ala Phe Asn Phe Gly Ser Tyr

Ile Asn Thr Gly Val Asp Pro Arg Thr Gly Gln Tyr Ser Ala Asn Ile

Asn	Ile	Ile 35	Thr	Leu	Arg	Pro	Asn 40	Asn	Val	Gly	Asn	<i>S</i> er 45	Glu	Gln	Thr
Leu	Ser 50	Leu	Ser	Phe	Ser	Pro 55	Leu	Thr	Thr	Leu	Asn 60	Asn	Gly	Phe	Gly
Ile 65	Gly	Trp	Arg	Phe	Ser 70	Leu	Thr	Thr	Leu	Asp 75	Ile	Lys	Thr	Leu	Thr 80
Phe	Ser	Arg	Ala	Asn 85	Gly	Glu	Gln	Phe	Lys 90	Cys	Lys	Pro	Leu	Pro 95	Pro
Asn	Asn	Asn	Asp 100	Leu	Ser	Phe	Lys	Asp 105	Lys	Lys	Leu	Lys	Asp 110	Leu	Arg
Val	Tyr	Lys 115	Leu	Asp	Ser	Asn	Thr 120	Phe	Tyr	Val	Tyr	Asn 125	Lys	Asn	Gly
Ile	Ile 130	Glu	Ile	Leu	Lys	Arg 135	Ile	Gly	Ser	Ser	Asp 140	Ile	Ala	Lys	Thr
Val 145	Ala	Leu	Glu	Phe	Pro 150	Asp	Gly	Glu	Ala	Phe 155	Asp	Leu	Ile	Tyr	Asn 160
Ser	Arg	Phe	Ala	Leu 165	Ser	Glu	Ile	Lys	Тут 170	Arg	Val	Thr	Gly	Lys 175	Thr
Tyr	Leu	Lys	Leu 180	Asn	Tyr	Ser	Gly	Asn 185	Asn	Cys	Thr	Ser	Val 190	Glu	Tyr
Pro	Asp	Asp 195		Asn	Ile	Ser	Ala 200	_	Ile	Ala	Phe	Asp 205	_	Arg	Asn
Asp	Tyr 210	Leu	Ile	Thr	Val	Thr 215		Pro	Тух	Asp	Ala 220	Ser	Gly	Pro	Ile
Asp 225		Ala	Arg	Phe	Lys 230	Met	Thr	Tyr	Gln	Thr 235		Lys	Gly	Val	Phe 240
Pro	Val	Ile	Ser	Thr 245	Phe	Arg	Thr	Pro	Thr 250		Tyr	Val	Glu	Leu 255	
Ser	Tyr	Lys	Glu 260		Gly	His	Lys	Val 265		Asp	Thr	Glu	Туг 270		Pro
Tyr	Ala	Ala 275		Leu	Thr	Il∈	280		Gly	/ Asr	Gly	Glr. 285		Ala	Val
Ser	Lys 290		Tyr	Glu	Tyr	Ser 295		· Val	. His	asr	Phe 300		ı Gly	Tyr	Ser
Ser 305		Arg	Thr	Ser	Phe 310	-	Ser	Ser	Glr	Asp 315		Lev	ı Tyr	Leu	Val 320
Thr	Gly	Lys	Tyr	Thr. 325	Tyr	Ser	Ser	: Ile	Glu 330		y Val	. Leu	ı Asp	Gly 335	
Ser	· Val	. Val	Ser 340		. Ile	: Glu	ı Arç	y Val 345		e Asr	ı Lys	: Phe	His 350		. Met
Thr	Lys	355		a Lys	Thr	Glr	360	-	ı Lys	s Arg	j Ile	Th: 365	_	Glu	ı Ile
Thu	Tyr 370		n Glu	ı Asp	Leu	375		s Sei	r Phe	e Sei	c Glu 380		n Pro	o Glu	ı Asr

Leu Gln Gln Pro Ser Arg Val Leu Thr Arg Tyr Thr Asp Ile Gln Thr Asn Thr Ser Arg Glu Glu Thr Val Asn Ile Lys Ser Asp Asp Trp Gly Asn Thr Leu Leu Ile Thr Glu Thr Ser Gly Ile Gln Lys Glu Tyr Val Tyr Tyr Pro Val Asn Gly Glu Gly Asn Ser Cys Pro Ala Asp Pro Leu Gly Phe Ser Arg Phe Leu Lys Ser Val Thr Gln Lys Gly Ser Pro Asp Ala Ala Gln Ser Val Ala Asn Lys Val Ile His Tyr Thr Tyr Gln Lys Phe Pro Thr Phe Thr Gly Ala Tyr Val Lys Glu Tyr Val Ser Lys Val Ser Glu Thr Ile Asp Asn Lys Ile Ala Arg Thr Phe Ser Tyr Val Asn 505 Ser Pro Thr Ser Lys Ser His Gly Ser Leu Ala Lys Ile Thr Ser Val Met Asn Asn Gln Gln Thr Val Thr Thr Phe Lys Tyr Glu Tyr Ser Glu Ser Glu Met Thr Thr Asn Ala Thr Val Thr Gly Phe Asp Gly Ala His 550 Met Glu Ser Lys Asn Val Thr Ser Ile Tyr Thr His Arg Gln Leu Arg 565 570 575 Lys Val Asp Val Asn His Val Ile Thr Asp Gln Ser Tyr Asp Leu Leu Gly Arg Ile Thr Gly Gln Ile Ile Asp Pro Gly Thr Ala Arg Glu Ile Lys Arg Asn Tyr Val Tyr Gln Tyr Pro Gly Gly Asp Glu Asn Asp Phe 615 Trp Pro Val Met Ile Glu Val Asp Ser Gln Gly Val Arg Arg Lys Thr His Tyr Asp Gly Met Gly Arg Ile Cys Ser Ile Glu Glu Gln Asp Asp Asp Gly Ala Trp Gly Thr Ser Gly Ile Tyr Gln Gly Thr Tyr Arg Lys Val Leu Ala Arg Gln Tyr Asp Val Leu Gly Gln Leu Ser Lys Glu Ile Ser Asn Asp Trp Leu Trp Asn Leu Ser Ala Asn Pro Leu Val Arg Leu 695 Ala Thr Pro Leu Val Thr Thr Lys Thr Tyr Lys Tyr Asp Gly Trp Gly Asn Leu Tyr Ser Thr Glu Tyr Ser Asp Gly Arg Ile Glu Leu Glu Ile

				725					730					735	
His	Asp	Pro	Ile 740	Thr	Arg	Thr	Ile	Thr 745	Gln	Gly	Val	Lys	Gly 750	Leu	Gly
Met	Leu	Asn 755	Ile	Gln	Gln	Asn	Asn 760	Phe	Glu	Gln	Pro	Ala 765	Ser	Ile	Lys
Ala	Val 770	Tyr	Pro	Asp	Gly	Thr 775	Ile	Tyr	Ser	Thr	Arg 780	Thr	Tyr	Arg	Tyr
Asp 785	Gly	Phe	Gly	Arg	Thr 790	Val	Thr	Glu	Thr	Asp 795	Ala	Glu	Gly	His	Ala 800
Thr	Gln	Ile	Gly	Tyr 805	Asp	Val	Phe	Asp	Arg 810	Ile	Val	Lys	Lys	Thr 815	Leu
Pro	Asp	Gly	Thr 820	Ile	Leu	Glu	Ser	Ala 825	Tyr	Ala	Ser	Phe	Ser 830	His	Glu
Glu	Leu	Ile 835	Ser	Ala	Leu	Asn	Val 840	Asn	Gly	Thr	Gln	Leu 845	Gly	Ala	Leu
Val	Tyr 850	Asp	Gly	Leu	Gly	Arg 855	Val	Ile	Ser	Asp	Thr 860	Val	Gly	Gly	Arg
Lys 865	Thr	Glu	Tyr	Leu	Tyr 870	Gly	Pro	Gln	Gly	Asp 875	Lys	Pro	Ile	Gln	Ser 880
Ile	Thr	Pro	Ser	His 885	Asn	Lys	Gln	Asn	Met 890	Asp	Tyr	Leu	Tyr	Tyr 895	Leu
Gly	Ser	Val	Met 900	Ser	Lys	Phe	Thr	Thr 905	Gly	Thr	Asp	Gln	Gln 910	Asn	Phe
Arg	Tyr	His 915	Ser	Lys	Thr	Gly	Thr 920	Leu	Leu	Ser	Ala	Ser 925	Glu	Gly	Val
Ser	Gln 930	Thr	Asn	Tyr	Ser	Tyr 935	Phe	Pro	Ser	Gly	Val 9 4 0	Leu	Gln	Arg	Glu
Ser 945	Phe	Leu	Arg	Asp	Asn 950	Lys	Pro	Ile	Ser	Ser 955	Gly	Glu	Tyr	Leu	Тут 960
Thr	Met	Ser	Gly	Leu 965	Ile	Gln	Arg	His	Lys 970	Asp	Ser	Phe	Gly	His 975	Asn
His	Val	Tyr	Ser 980	Tyr	Asp	Ala	Gln	Gly 985		Leu	Val	Lys	Thr 990	Glu	Gln
Asp	Ala	Gln 995	Tyr	Ala	Thr		Glu 1000		Asp	Asn		Gly 1005	Arg	Leu	Ile
Thr	Thr 1010	Thr	Thr	Lys	Asp	Thr 1015	Thr	Ser	Leu		Gln 1020	Leu	Val	Thr	Lys
Ile 025	Glu	Tyr	Asp		Phe 1030	Asp	Arg	Glu		Lys 1035	Arg	Ser	Leu		Ser 1040
Asp	Phe	Ser		Gln 1045	Val	Ile	Thr		Ser 1050	Tyr	Thr	Lys		Asn 1055	Gln
Ile	Ser		Arg 1060		Thr	Ser		Asp 1065	Gly	Val	Val		Lys 1070	Asn	Glu

- Arg Tyr Gln Tyr Asp Asn Asn Gln Arg Leu Ser Gln Tyr Gln Cys Glu 1075 1080 1085
- Gly Glu Gln Ser Pro Ile Asp His Thr Gly Arg Val Leu Asn Gln Gln 1090 1095 1100
- Ile Tyr His Tyr Asp Gln Trp Gly Asn Ile Lys Arg Leu Asp Asn Thr 105 1110 1115 1120
- Tyr Arg Asp Gly Lys Glu Thr Val Asp Tyr His Phe Ser Gln Ala Asp 1125 1130 1135
- Pro Thr Gln Leu Ile Arg Ile Thr Ser Asp Lys Gln Gln Ile Glu Leu 1140 1145 1150
- Ser Tyr Asp Ala Asn Gly Asn Leu Thr Arg Asp Glu Lys Gly Gln Thr 1155 1160 1165
- Leu Ile Tyr Asp Gln Asn Asn Arg Leu Val Gln Val Lys Asp Arg Leu 1170 1180
- Gly Asn Leu Val Cys Ser Tyr Gln Tyr Asp Ala Leu Asn Lys Leu Thr 185 1190 1195 1200
- Ala Gln Val Leu Ala Asn Gly Thr Val Asn Arg Gln His Tyr Ala Ser 1205 1210 1215
- Gly Lys Val Thr Asn Ile Gln Leu Gly Asp Glu Ala Ile Thr Trp Leu 1220 1225 1230
- Ser Ser Asp Lys Gln Arg Ile Gly His Gln Ser Ala Lys Asn Gly Gln 1235 1240 1245
- Ser Val Tyr Tyr Gln Tyr Gly Ile Asp His Asn Ser Thr Val Ile Ala 1250 1255 1260
- Ser Gln Asn Glu Asn Glu Leu Met Ala Leu Ser Tyr Thr Pro Tyr Gly 265 1270 1275 1280
- Phe Arg Ser Leu Ile Ser Ser Leu Pro Gly Leu Asn Gly Ala Gln Val 1285 1290 1295
- Asp Pro Val Thr Gly Trp Tyr Phe Leu Gly Asn Gly Tyr Arg Val Phe 1300 1305 1310
- Asn Pro Val Leu Met Arg Phe His Ser Pro Asp Ser Trp Ser Pro Phe 1315 1320 1325
- Gly Arg Gly Gly Ile Asn Pro Tyr Thr Tyr Cys Gln Gly Asp Pro Ile 1330 1335 1340
- Asn Arg Ile Asp Leu Asn Gly His Leu Ser Ala Gly Gly Ile Leu Gly 345 1350 1355 1360
- Ile Val Leu Gly Ala Ile Gly Ile Ile Val Gly Ile Val Ser Leu Gly 1365 1370 1375
- Ala Gly Ala Ala Ile Ser Ala Gly Leu Ile Ala Ala Gly Gly Ala Leu 1380 1385 1390
- Gly Ala Ile Ala Ser Thr Ser Ala Leu Ala Val Thr Ala Thr Val Ile 1395 1400 1405
- Gly Leu Ala Ala Asp Ser Ile Gly Ile Ala Ser Ala Ala Leu Ser Glu 1410 1415 1420

Lys Asp Pro Lys Thr Ser Gly Ile Leu Asn Trp Ile Ser Ala Gly Leu 425 1430 1435 1440	
Gly Val Leu Ser Phe Gly Ile Ser Ala Ile Thr Phe Thr Ser Ser Leu 1445 1450 1455	
Val Lys Ser Ala Arg Ser Gly Ser Gln Ala Val Ser Ala Gly Val Ile 1460 1465 1470	
Gly Ser Val Pro Leu Glu Phe Gly Glu Val Ala Ser Arg Ser Ser Arg 1475 1480 1485	
Arg Trp Asp Ile Ala Leu Ser Ser Ile Ser Leu Gly Ala Asn Ala Ala 1490 1495 1500	
Ser Leu Ser Thr Gly Ile Ala Ala Ala Ala Val Ala Asp Ser Asn Ala 505 1510 1515 1520	
Asn Ala Ala Asn Ile Leu Gly Trp Val Ser Phe Gly Phe Gly Ala Val 1525 1530 1535	
Ser Thr Thr Ser Gly Ile Ile Glu Leu Thr Arg Thr Ala Tyr Ala Val 1540 1545 1550	
Asn His Gln Thr Trp Glu Leu Ser Ser Ser Ala Gly Thr Ser Glu Glu 1555 1560 1565	
Val Lys Pro Ile Arg Cys Leu Val Ser His Arg Trp Asn Gln Lys Gln 1570 1575 1580	
<210> 7 <211> 18 <212> DNA <213> Artificial Sequence <220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 7 acacagcagg ttcgtcag	18
<210> 8 <211> 18 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 8 ggcagaagca ctcaactc	18
<210> 9 <211> 20 <212> DNA <213> Artificial Sequence	
<211> 20 <212> DNA	

- 22 -

```
20
attgatagca cgcggcgacc
<210> 10
<211> 22
<212> DNA
<213> Artificial Sequence
<220>
<223> Description of Artificial Sequence:oligonucleotide
<400> 10
                                                                   22
ttgtaacgtg gagccgaact gg
<210> 11
<211> 37948
<212> DNA
<213> Photorhabdus luminescens
<220>
<221> CDS
<222> (15171)..(18035)
<223> orf5
<220>
<221> CDS
<222> (23768)..(31336)
<223> hph2
<220>
<221> CDS
<222> (31393)..(35838)
<223> orf2
<400> 11
tgttgctgga ccgtggagat tatgcctatc gtcagttaga acgagacacg ctcaatgaag 60
ccaagatgtg gtatatgcaa gcactgcatc tgttaggcga taaacctcat ctatcgttca 120
gttcagagtg gagcaaaccg agtttaggcg acgctgccgg aacggaaaga caaaagcaac 180
acgcccaagc aatggccgct ctgcgacaag gtgatgttag tcggcacaac aacccgacag 240
 atcttttctt gccacaggtc aatgaagtga tgcaaaacta ttggcaaaaa ttggaacaac 300
 ggetgtataa cetgegteat aaceteacta ttgaeggeea aeegetacat etgeetattt 360
 acgetacace ggcagateca aaageattae ttagegeege tgtegetage teegaaggtg 420
 gggtagetet eteacageea titatgteae tgtggegtit eecacacatg etggaaaaeg 480
 cgcgtggtat ggtcagtcag ctcactcaat tcggctctac gctacaaaat attatcgaac 540
 gtcaggatgc ggaagcttta aacacgctct tgcagaatca agcagcagaa ctgatattga 600
 ctcatctcag catacaggac aaaaccatcg cagagctgga tgcggaaaaa atcgtactgg 660
 aaaaatccaa agccggggcg caatcacgct ttgacagcta caaaaagtta tacgacgaaa 720
 atatcaatgc gggtgaaaac cgggctatag cattgcatgc ctccgttgct ggcctcagca 780
 ctgccctgca agcatcacgt ctggcgggcg ctgcgcttga tctggcgccc aacattttcg 840
 gtctcgctga tggcggtagc cgttggggag cgattgccga agcgacaggt aatgttatgg 900
```

aatteteege	cagtgtgatg	aacaccgaag	cggataaaat	cagccagtca	gaagcctatc	960
gccgtcgccg	tcaggaatgg	gaaatccagc	gtaatcatgc	cgaagcagag	ataaaacaga	1020
tcgatgctca	acttcaatca	ctggcagtac	gccgtgaagc	cgcggtattg	cagaaagcca	1080
gcctaaaaac	ccaacaggaa	cagactcatg	ctcaattgac	tttcctgcaa	cgtaaattca	1140
gtaatcaagc	gttgtactac	tggctacgcg	gtcggctagc	tgctatttac	ttccaatttt	1200
acgatttggc	cgtagcgcgt	tgtctgatgg	ctgaaatggc	ttatcgttgg	gagactaatg	1260
agaccgcggc	aagctttatc	aaacccggcg	cctggcaggg	aacccatgcg	ggtttactgg	1320
ctggtgaaac	cttgatgctg	aatctggcgc	aaatggaaga	tgcccatttg	aggtgggatc	1380
aacgcgctct	ggaagtggaa	cggaccattt	cattgacgca	acactatgga	gcactgccag	1440
aaaaatcgtt	taatttagcc	acacggattt	ctaccctgct	agcaggtggt	acaactgact	1500
ccattgatga	tcatcccgtt	acattagaaa	acgaccaact	tagtgccaaa	atctctctgt	1 560
caggtctgtc	attagataat	gactacccag	atggcaacgg	cgtaggcaac	attegaegea	1620
ttaaacaaat	cagtgtcacc	ttgccagccc	tgttaggacc	atatcaggat	gtacaagcta	1680
ttctgtccta	cggaggaagt	gaaatcggat	tagctgaaag	ctgtaaatca	ctggcgatct	1740
ctcatgggat	caatgacagt	ggtcaattcc	agttggattt	taacaatggt	aagttcctgc	1800
cgtttgaagg	gattgcgatt	aacgatactg	gcacattgac	actcagtttc	ccccaatgcg	1860
actgtcaaac	aagaaaacat	gttgcagact	ttgagtgata	ttattctgca	tattcgctat	1920
accatccgcc	aataaccacc	tcaattaaat	ассаааааса	ggctcctaaa	cggggcctga	1980
acttttcacg	aatatatacc	actcacagtc	tgctctcttt	acctgtctga	cgctcgttat	2040
aacagagata	tttccttttc	tcgtgagtcc	catcacctac	tataaaatat	caaccctctt	2100
ctttttcata	atatgcaata	tgtaacaaat	gcaattattt	catttagtta	ttgttaacta	2160
gttatattac	ttatgatgta	attataaatt	ttgttattgc	atcacaatag	ccatttaaat	2220
aaataataac	gttgtgaaat	agttgatagt	taaatggtgt	ttttatttag	ccgttatttt	2280
caacccaatt	tcagaccgct	atcagacgtt	acctgtgttg	cctttgtttt	gatagatata	2340
aataacctta	tttatatcca	cggtactcag	accagcataa	atgttttatt	tacctaacat	2400
ttaaaaggaa	taaacatgaa	cacactcaaa	tccgaatatg	aaaacgcgtt	agtagcaggt	2460
tttaataatc	taaccgatat	ttgtcatctc	tcttttgacg	aacttcgcaa	aaaagtgaag	2520
gacaaactct	catggtcaca	gacccaaagc	ttatatcttg	aagcacagca	agtgcaaaag	2580
gacaatctcc	tgcatgaagc	ccatattctg	aaacgcgcca	atcctcattt	acaaagtgcg	2640
gtccatcttg	ccctgacaac	acctcatgct	gaccagcaag	gttataatag	cagatttggc	2700
aatcgcgcca	gcaaatatac	agccccaggc	gcaatttctt	ccatgttttc	tcctgcggct	2760
tatttagctg	aactttatcg	tcaggcacgg	aatttacatg	atgaaaattc	tatttatcat	2820

PCT/EP99/01015

ttggatacac ggcgtccgga tctaaaatca ttggtgctca gccagaaaaa tatggatacg 2880 gagatttcca cactttctct gtctaatgac atgttgctag agggtattaa aactctgttc 2940 aaggacaagc tgctggaggc tctgaagaat attaaatctc tgtccaagga cgagctgctg 3000 gaggetttga agaatattaa acetetgtee aaggaegate tgetggagge tttgaagaat 3060 attaaacctc tgtccaagga cgatctgctg gaggctttga agaatattaa acctctgtcc 3120 aaggacgatc tgctggaggc tttgaagaat attaaacctc tgtccaagga cgatctgctg 3180 gaggetttga agaatattaa acetetgtee aaggaegate tgetggagge tttgaagaat 3240 attaaacctc tgtccaagga cgatctgctg gaggctttga agaatattaa acctctgtcc 3300 aaggacgatc tacaggaatg tattgaaatt ctattcaatc tggacagcca cactaaagta 3360 atqaaaqcgt tatccaattt ccgcgtttct ggcatgatgc catatcacga tgcttatgaa 3420 agcqtqcqta aggttgttca attacaggct ccggtgtttg aacacgttgt tagtacatca 3480 ctagaaacga ctatcgatga actaaaatat caagcttctt tgttggaaat taattcttct 3540 qtctcqccta aattatttac tatcttgact gaagaaatta ctacaatcaa tgcaagaagt 3600 ctctatgagg aaaattttgg taatattaaa ccttctctaa taggaaaacc ggaatatctg 3660 aaaagttatt acaatctgag tgatgaagag tttagcgatt tcattaaaat aagaactata 3720 cttcttccaq aaqaaqaaat agcaattact gatcttgcat cgcgtactac tagtacacaa 3780 cagactateg aaaateetga ttategtget etattgaaaa ttaataagtt tattegteta 3840 ttcaaagcta taaacttatc accgacggta ttaagtggaa tcctccgcag catcagcaca 3900 qaattcaata tcaataaaga aatattacaa aaaatctttc gtgttaaata ctatatgcaa 3960 cqttatqqta ttgacactga gactgcatta atactatgca aggtaccaat ttcacaatat 4020 atcaatgacg gacatctaag tcagtttgat cgtttattta attcccccaa actgaatggt 4080 caagattttt ccgtcaatgg tactcagaat attgatttaa ccctaagcag taccaacaac 4140 tggaataaaa cagtacttaa acgtgctttt aacctcgatg atatctcatt aaatcgacta 4200 ctaaaaatta ccaatccggt caatactacc gaaatgataa ctaatgatat agagaatctt 4260 teteatetet ataggacaaa attaetggca gatateeate agttaaetat tgatgaaetg 4320 gggttactgt tggaagccat aggtaaagga acaaccaatt tatctgagat tactcctgac 4380 aatctggtta ctctaattaa caaactctat gctgtcacta gctggctacg tacacaaaag 4440 tggagtgtct atcagttgtt tatgatgact actgataaat ataacaaaac cctaaccccg 4500 gaaatacgga atttactgga taccgtctac aatggcttgc aagattttaa taagaagatg 4560 ttgaaagctg aagaagatct agagaaaacc aaaaagaaat tgcagagcgc caaggaaaat 4620 ctggaaaaat tcccggaaaa ccagccacaa ctccaagaag acaggaaaaa agcccagaga 4680 agactgaata aagctgaaga gacccacgaa aaagccgaga aaaacctaga tgaggtcagg 4740 aaaaatctgc caaaagccat atctccttat atcgccgccg ctctgcaatt accatctgaa 4800

catgcggcat	attccatact	catctgggca	gataatctgg	aacccggcat	aggaaaaatg	4860
acagcggaaa	aattatggaa	ctggttgcgg	aaaaatcccg	ttacggctca	acctgaattc	4920
caaaaacaag	ctgaacctgt	ggtccagtat	tgccagcgcc	tggcacaact	agcgttgatt	4980
taccgttcta	ccggccttaa	cgaaaacacc	ttaagtctgt	ttgtgacaaa	gccgcaacac	5040
tttgttatta	aaaccaaagc	acccgaaaca	actgaaacaa	caccagcaca	tgacgtatca	5100
acactaatgt	cactaacgcg	ttttactgac	tgggttaact	cactaggtga	aaacgcctct	5160
tctgtactaa	ccgaatttaa	aaaaggaaca	ttaacagcag	aactattggc	taaggctatg	5220
aatcttgata	aaaatctact	ggagcaagcc	aagattcagg	caaaaactga	tttgtccaac	5280
tggccatcta	tcgacaacct	attgcagtgg	attaacatct	cacgtcaatt	gaacatctct	5340
ccacaaggca	tttccacact	gactcaagta	ttgaccgcag	aacctcccgc	taactatacc	5400
caatgggaaa	acgccgctgc	gatattaacc	gccgggctgg	acacccaaaa	gactaacgcc	5460
ctacatgcgt	ttctggatga	gtctcgcagt	gctgcgttaa	gcacatacta	tatttattct	5520
cataaccaaa	aagatcgaga	agcaagaaaa	catacagtaa	ttaaaaaccg	tgatgatcta	5580
tatcaatacc	tattgatcga	taaccaagtt	tccgccgaca	ttaaaactac	agagatcgct	5640
gaagctatcg	ctagtateca	actgtatatt	aaccgcgcgt	tgaaaaatat	ggagggagat	5700
actgtcacaa	gtgtcaccag	ccgctcattc	ttcaccaact	gggataaata	caataaacgc	5760
tacagcactt	gggccggtat	ggctaaactc	ctttactatc	cagagaatta	catcgatccg	5820
acgctacgta	ttgggcagac	aaaaatgatg	gatacgttgc	tgcaatccat	cagccaaagc	5880
caattaaata	tcgataccgt	agaagatgcc	tttaaatctt	acctaacatc	attcgaacag	5940
gtggctaatc	tggaaateet	cagcgcctac	catgacaaca	ttaataatga	tcaaggatta	6000
acttacttta	tcggacgtag	taaaacagaa	gtgaatcaat	attattggcg	cagtgtggat	6060
cacaataaat	ccagcgaagg	taaattcccc	gctaatgcct	ggagtgagtg	gtacaaaatt	6120
gattgtccaa	ttaaccccta	caaagatact	attcacccgg	taattttcca	atctcgcctg	6180
tatcttatct	ggctggaaca	aaaaaaggcg	actaaacagg	aaggtgataa	aaccgcctcg	6240
ggttattatt	atgaactgaa	attagcgcat	atccgttatg	acggcacctg	gaatacacca	6300
gtcacctttg	atgtaaacca	aaaaatatcc	gatttaaatc	tgggaaataa	aacacctgga	6360
ctttactgct	caagctttca	aggcagagat	gaattgctgg	tgatgtttta	taaaaaacaa	6420
gatcaattaa	atcaatacac	aaacacagta	ccaataaaag	gactatatat	cacttccaat	6480
atgtcttcta	aggaaatgac	acctgaaaat	cacaaaccta	acgcttataa	acagtttgat	6540
actaatagta	ttattggtgt	caataatcgc	tatgcagaaa	gctacgaaat	cccttcatca	6600
gtaaatagta	ataacggtta	tgattgggga	gatggctatc	tgagtatggt	gtatggcgga	6660
aatatttcag	ccatcaaact	. ggagtcctca	tcagataagt	taaaactctc	accaaggtta	6720

	ataatggact	tataggccaa	caacgcaacc	aatgcaacct	gatgaagaaa	6780
_						
	ttggtgataa					
	aaaaattcat					
	gtctgctgtt					
	cctctggctc					
aactatggtt	atattgataa	caaacatagt	gacgatgttc	catttaaaca	atatttctat	7080
atggatgacc	acggtggtat	tgacactgat	gtttcaggga	tattatctat	taatacgaac	7140
attaatcatt	caaaagttaa	agtaatagtg	aaagccgaag	gtatcacaga	gcaaactttt	7200
gtagcgagcg	aaaacagtaa	tgtccccacc	aatccgtccc	gcttcgaaga	aatgaattat	7260
cagtttaaag	agcttgaaat	agatatetee	acactgacat	ttcataataa	tgaagcaagt	7320
attgatatca	cctttatcgc	atttgctgag	aaatttgacg	ataatagtaa	tgatcgtaac	7380
ttaggcgaag	aacatttcag	tattcgtatt	atcaaaaaag	cggaaactga	taatgccctg	7440
accetgeace	ataatgcaaa	cggggcgcaa	tatatgcagt	ggggaaactc	ttgtattcgc	7500
cttaatacgc	tatttgcccg	tcaattaatt	agccgagcca	acgcggggat	agatactatt	7560
ttgagcatgg	acactcagaa	tattcaggaa	cctaaattag	gagaagatto	tcctgatgct	7620
atggaaccaa	tggacttcaa	cggcgccaac	agcctctatt	tctgggaact	gttctactac	7680
accccgatgo	tgattgctca	acgtttgctg	cacgaacaaa	acttcgatga	ggctaaccgt	7740
tggctgaaat	atgtctggaa	cccatccggt	tatattgtca	atggtcaaat	gcaacattac	: 7800
cgctggaatg	ttcgcccatt	acaagaagac	actagttgga	acgatgatco	gttggattca	7860
tttgatcctg	g ataccatago	tcaacatgat	ccaatgcact	acaaagtcgc	cacctttato	7920
cgcaccctag	g atctgttgat	. cgaacgggga	gattacgcct	atcgccaatt	ggagcgggad	7980
acactcgctg	g aagccaaaat	gtggtatatg	caggcactgo	atctattggg	g tgataaacct	8040
catctacca	c tcagttcagc	: atggaatgat	ccagagctag	g aagaggccgo	agctcttgaa	a 8100
aaacaacag	g cacatgccaa	agaaatagca	gatttacgad	c aaggacttco	tacatccaca	a 8160
gggtctaaa	g atgaaatcaa	a aacagatctt	tteetgeege	c aagtcaacga	a agtgatgct	g 8220
agctactgg	c agaaactaga	a acaacggttg	tataacctg	gccataacc	t ctctattga	8280
ggtcaacct	t tacatttgco	tattttcgca	a acaccagcag	g atccaaaag	c gctgctcage	2 8340
geegetgte	g ccagttcaca	a aggtggaagt	aatcttcca	t cagaattta	t atcagtgtg	g 8400
	c atatgctgga					
	c aaaatatta	*				
	g cagaactga					
	g aaaaaactg					
	a aattctacg					
		5 5	3 223			

cgtgcttccg	tcgctggcat	ctctacagcc	cttcaagcat	cacatctggc	gggcgcagca	8760
cttgatctgg	cgcccaacat	cttcggcttc	gctgatggtg	gcagccgttg	gggggcgatc	8820
gcccaagcca	caggtaatgt	catggagttc	tccgccagtg	ttatgaacac	cgaagcggat	8880
aaaatcagcc	aatctgaagc	ctaccgtcgg	cgtcgtcagg	aatgggaaat	tcagcgtaat	8940
aacgccgagg	cagagctgaa	acaaatcgat	gctcaacttg	gttcgctggc	agtgcgccgt	9000
gaagccgcag	tattgcagaa	aaccagccta	aaaacccaac	aagagcagac	tcatgcacaa	9060
ctgaccttcc	tgcaacataa	gttcagtaat	caggcgctgt	acaactggct	gcgtggtcga	9120
ttgtccgcca	tttacttcca	gttctatgat	ttaacggtag	ctcgctgttt	gatggcggaa	9180
atggcctatc	gctgggagac	taacgatacc	gcatcacgct	ttatcaaacc	cggcgcctgg	9240
cagggaaccc	atgccggttt	gctcgcgggt	gaaaccttaa	tgctgaatct	ggcacagatg	9300
gaagatgccc	acctgaaaca	ggataaacgc	gtactggagg	tagaacgtac	cgtttcgctg	9360
gccgaagtct	atgccaaatt	accgcaagat	aaatttatcc	tgactcagga	aatagagaag	9420
ttggtgagta	aaggttcagg	cagggccggc	aaggacaata	ataagctggc	gtttagtacc	9480
aataccaata	cctctctaga	agcgtccatt	tcgttatcta	ccttgaacat	tagcagcgat	9540
tatcctgatt	ctattggtaa	aacccgtcgt	attaaacaga	tcagcgttac	cctgccagca	9600
ctgctaggac	cctatcagga	tgtgcaagca	attctgtctt	acagcggaaa	agcctctgaa	9660
ttggctgaaa	gttgcaaatc	attagcggtt	tctcatggga	tgaatgacag	cggtcagttc	9720
caactggatt	tcaacgatgg	caaattcctg	ccgttcgaag	gaatcaaaat	cgatgaaggt	9780
acgctgacat	tgagcttccc	aaatgcaatt	agtaaagaag	acaaaaaaga	cgaaaaaggc	9840
aaacaacaag	ccatgctgga	gagtetgaac	gacatcattc	tgcatattcg	ctacaccatt	9900
cgccaataac	gattttaatt	aagtgctaaa	acaggcccct	aagcggggcc	tgcaaggagt	9960
ctttcatgca	aaattcacaa	gatttcagta	ttacagaact	atcattgccc	aaaggaggag	10020
gcgctatcac	gggaatgggg	gaagctttaa	ccccaccgg	gccggatggg	atggccgcgc	10080
tgtctctgcc	gttgcctatc	tctgccgggc	gcggttatgc	tccgtcactc	gccttaaact	10140
acaacagcgg	cgccggtaac	agcccatttg	gtctgggctg	ggattgcaac	gttatgacca	10200
teegeegeeg	cacccatttt	ggcgttccac	attatgatga	aaccgatacc	tttctggggc	10260
cagatggcga	ggtactggtg	gtagcggatc	aatcccgcga	cgaatcgaca	ttacagggta	10320
					agtcatttca	
					acagattttt	
					caagcccgga	
					tctgtgtcac	
					tgcgaagctg	
				_		

atgaaattac tetecateca	caggccgccg	cgcaacgtta	tctacacaca	gtgtattacg	10680
gcaaceggac agccagcaaa	acgttacccg	gtctggatgg	cagegeeeca	ccacaagcag	10740
actggttatt ctatctggta	tttgattacg	gcgaacgcag	taacaacctg	agaacgccgc	10800
cagcattttc gactacaggt	agctggcttt	gtcgccagga	ccgtttttcc	cgttatgaat	10860
atggttttga gattcgtacc	cgccgcttat	gccgtcaggt	attgatgtat	caccacctgc	10920
aagetetgga tagegagata	aaagaacaca	acggaccaac	gctggtttca	cgcctgatac	10980
tcaattatga cgaaagcgca	atcgccagca	cgctggtatt	cgttcgtcga	gtaggccacg	11040
agcaagacgg tactgccgtc	accetgeege	cattagaatt	ggcgtatcag	gatttttcac	11100
cgcaacataa cgctcgctgg	caatcgatga	atgtgctggc	aaacttcaat	gccattcagc	11160
gctggcaact agttgatcta	aaaggcgaag	gattccccgg	tctgctatat	caagataaag	11220
gegeetggtg gtaeegetee	gcacaacgtt	ttggcaaaat	tggctcagat	gccgtcactt	11280
gggaaaaaat gcaacctttg	tcggttatcc	cttccttgca	aagtaatgcc	tcgctggtgg	11340
atatcaatgg agacggccaa	cttgactggg	ttatcaccgg	accgggatta	cggggatatc	11400
atagtcagca tccagatggc	agttggacac	gttttacccc	gctcaacgct	ctgccagtgg	11460
aatatactca tccacgcgcg	caactcgccg	atttaatggg	agctggactt	tctgatttag	11520
tactgatcgg ccctaagagt	gtacgtttat	atgccaatac	ccgcgacggc	tttgccaaag	11580
gaaaagatgt agtgcaatcc	ggtgatatca	cactgccagt	accgggcgcc	gatccgtgta	11640
agttggtggc atttagtgat	gtattgggtt	ccggtcaggc	acatctggtt	gaagtgagcg	11700
cgactaaagt cacctgctgg	cctaatctgg	ggcacggacg	ttttggtcaa	ccaattactc	11760
ttccgggatt tagccaacca	gaagcgacgt	ttaatcctgc	tcaagtttat	ctggccgatc	11820
tagatggcag cggcccgact	gatctgattt	atgttcacac	agatcgtctg	gatatcttcc	11880
tgaataaaag cggcaacggc	ttcgccgcac	cagtaactct	ccccttccca	gccggagtgc	11940
gttttgatca tacctgtcag	ttacaagtgg	ccgatgtaca	agggttaggc	gtcgccagcc	12000
tgatattaag tgtgccgcat	atgactcccc	atcactggcg	ttgcgatctg	accaacacaa	12060
aaccgtggtt actcagtgaa	atgaacaaca	atatgggggc	tcatcacacc	ctgcgttacc	12120
gtagttccgc ccagttctgg	ctggatgaaa	aagccacggc	actggatgcc	ggacaaatac	12180
cagtttgtta tctacccttc	ccggtacaca	ccctatggca	aacggaaata	gaggatgaaa	12240
tcagcggcaa caaattagtc	acaatactac	gttatgcaca	tggcgcctgg	gatggacgtg	12300
agcgagaatt tcgcggattt	ggttatgttg	aacagaaaga	cagccatcaa	ctggcccaag	12360
gcagtgcgcc agaatgcaca	ccacctgcac	tgacccaagg	caacgcgcct	gaactcacat	12420
caccegeget gacceaagge	aacgctccag	aactcacacc	acctgcgatg	acccaaagca	12480
acgcgcctga actcacatca	cccgcgctga	cccaaggcaa	cgcgccagaa	ttcacatcac	12540
ccgcgctggc ccaaggcaat	gcgccagaac	tcacaccacc	tgcgatgacc	aaaaactggt	12600

atgccaccgg	aatacccatg	atagataaca	cattatcgac	agagtattgg	catggtgatc	12660
accaagcttt	tgccggtttt	tcaccacgct	ttacgacctg	gcaagatggt	caagatattc	12720
tgctcacacc	ggaaaatgat	aacagtcagt	actggctaaa	ccgggcactg	aaaggtcaac	12780
tgctacgcag	tgaactgtac	ggcgaggatg	gcagtacaca	ggaaaaaatt	ccctacacag	12840
tcactgaatt	tegeceacag	gtacgtcggt	tacagcatac	cgatageega	taccttgtgc	12900
tttggtcatc	tgtagttgaa	agccgcaact	atcattacga	acgtatcgcc	agcgatcctc	12960
aatgcagcca	aaagattacg	ctatccagcg	atctatttgg	tcaaccgcta	aaacaggttt	13020
cggtacagta	tccacgccgc	cagcaaccgg	caagcagtcc	gtatcctgat	acgttgcctg	13080
ataagttatt	tgctaacagc	tatgatgacc	agcaacacaa	attacggctc	acctatcaac	13140
agttcagttg	gcatcatctg	accgacaata	ccattctgat	gttaggatta	ccggatagta	13200
cccgcagcga	tatctttgct	tatagcgctg	aacatgtccc	tactggtggt	ctaaatctgg	13260
aaatcctaaa	tgataaaaat	agtctgattg	cggagaataa	acctcgtgaa	tacctcggcc	13320
agcaaaaaac	cgtttatacc	gacgggcaaa	atgcaacgcc	atcgcaaacg	ccaacacgac	13380
aagcgctgat	tgccttcacc	gagacaacag	tatttaatca	atccacacta	tcagcgtttg	13440
atgggagtat	ctcatctgct	caattgtcaa	cgacgctgga	acaagccgga	taccagcaaa	13500
cagattatct	attcccgcgc	actggagaag	ataaagtctg	ggcagctcgt	cgtggctata	13560
ctgattacgg	cacagccgaa	cagttctggc	ggccgcaaaa	acagagcaac	actcaactca	13620
cgggcaaaat	cacgctcact	tgggatgcaa	actattgcgt	cgtcacacaa	acccgggatg	13680
cggctggact	gacaacctca	gccagatatg	attggcgttt	tctgaccccc	gttcaactca	13740
cggatatcaa	cgacaatcag	caccttacca	cgctggatgc	actgggccga	ccaatcacac	13800
tgcgcttttg	gggaaccgaa	aacggtaaga	tgactggtta	ttcttcaccg	gaaaaaatat	13860
cgttttctcc	accatctgat	gttgacgccg	cgattaagtt	aacaacgcca	atccctgtag	13920
cacagtgtca	ggtctacgca	cccgaaagct	ggatgcccat	attaaagaaa	accctcaata	13980
acctggcaga	gcaagagcgg	aaagagttat	ataacacccg	aatcatcacc	gaagacggac	14040
gcatctgtac	cctagctcac	cgccgctggg	taaaaagcca	aagtgcagtc	acccagccaa	14100
tcaatctgtc	aaacggcagt	ccccgtttac	cccctcatag	cctcacattg	actacggatc	14160
gttatgaccg	cgatcttaag	caacagatto	gtcaacaagt	agtattcagt	gatggctttg	14220
gccgtttact	gcaagcatct	gtacgacatg	aagcaggcga	agcctggcaa	cgtaaccaag	14280
acggcgctct	ggtgacaaaa	atggaagata	ccaaaacgcg	ctgggcggtt	acgggacgca	14340
ctgaatatga	caataaggga	caaccgatac	gcacctatca	accctatttc	ctcaacgact	14400
ggcaatacgt	cagtaatgac	agtgcccggc	ggacagaaga	agcctatgca	gatacccatg	14460
tctatgatcc	cattggtcga	gaaatcaagg	tcactaccgo	: aaaaggctgg	ttccgtcgaa	14520

ccttgttcac tccttggttt actgtcaatg aagatgaaaa tgacacagct actgaggtga 14	580
aggtaaagaa gaaagaatgt aaagaaggta aagaaggtaa agatgtaatt tgatcaatcc 14	640
cgcccggttg aagggcggga aacataacat aatatagagg tgaaacgtgt cattcataat 14	1700
gccgtcagat actcaactta tgagttggtt gatcattggt tttattgcgg cctggggcgg 14	1760
attagtaagg tacctcattg atatacaaaa caaacaatgt aaatggaatt ggatcaacgt 14	1820
actctgtcaa ctcattatct cctgttttac cggtatattg ggaggactgc tgagttttga 14	1880
aageggegge ageceetata tgaettttge gattgeeggg etatttggea eeaegggaag 14	1940
ttctggattg aactggatct ggcgtcgcct ttttatgcat tatcgcgatg atggaggaaa 15	000
gcaataaggc attcccactg ccgcaaaaac catctgtctc cggcagttaa accgggaaat 15	6060
tacctactac aactattgta agaaaacgaa tatatagaaa aactaacatg cagataaaaa 15	5120
ctgcgattgc agaacagatg acacacaacg ccccaacaac gaggtaaatc atg aaa 15 Met Lys 1	5176
aac atc gat cct aaa ctt tat caa aag acc cct gtc gtc aac atc tac Asn Ile Asp Pro Lys Leu Tyr Gln Lys Thr Pro Val Val Asn Ile Tyr 5 10 15	5224
gat aac cga ggt cta acg atc cgt aac atc gac ttt cac cgt acc acc Asp Asn Arg Gly Leu Thr Ile Arg Asn Ile Asp Phe His Arg Thr Thr 20 25 30	5272
gca aac ggc gat acc gat atc cgt att act cgc cat caa tat gac tcc 15 Ala Asn Gly Asp Thr Asp Ile Arg Ile Thr Arg His Gln Tyr Asp Ser 35 40 45 50	5320
ctt ggg cac cta agc caa agc acc gat ccg cgt cta tat gaa gcc aaa 15 Leu Gly His Leu Ser Gln Ser Thr Asp Pro Arg Leu Tyr Glu Ala Lys 55 60 65	5368
caa aaa tot aac tit oto tig oag tat gat tig acc ggt aat att tig 19 Gln Lys Ser Asn Phe Leu Trp Gln Tyr Asp Leu Thr Gly Asn Ile Leu 70 75 80	5416
tgt aca gaa agc gtc gat gct ggt cgc act gtc acc ttg aat gat att 19 Cys Thr Glu Ser Val Asp Ala Gly Arg Thr Val Thr Leu Asn Asp Ile 85 90 95	5464
gaa ggc cgt ccg cta ctg aca gta act gca aca ggt gtc ata caa acc Glu Gly Arg Pro Leu Leu Thr Val Thr Ala Thr Gly Val Ile Gln Thr 100 105 110	5512
cga caa tat gaa acg tct tcc cta ccc ggt cgt ctg ttg tct gtt acc Arg Gln Tyr Glu Thr Ser Ser Leu Pro Gly Arg Leu Leu Ser Val Thr 115 120 125 130	5560
gaa caa ata cca gaa aaa aca tcc cgt atc acc gaa cgc ctg att tgg 1990 1990 1990 1990 1990 1990 1990 19	5608
get gge aat age gaa gea gag aaa aac eat aat ett gee age eag tge 19 Ala Gly Asn Ser Glu Ala Glu Lys Asn His Asn Leu Ala Ser Gln Cys 150 155 160	5656
gtg cgc cac tat gac acg gcg gga gtc acc cga tta gag agt ttg tca 1	5704

Val i	Arg	His 165	Tyr	Asp	Thr	Ala	Gly 170	Val	Thr	Arg	Leu	Glu 175	Ser	Leu	Ser	
ctg : Leu '																15752
caa Gln 195																15800
ctg Leu																15848
gct Ala																15896
tat Tyr								ggg Gly								15944
								tcc Ser								15992
								aat Asn			Ile					16040
					Gln					Thr					ccg Pro	16088
	_			Val	_		-		Arg				_	Pro	gta Val	16136
			Ile					Asp					Arg		tgg Trp	16184
		Glr					Glu					Tyr			ttg Leu	16232
tat Tyr 355	Glr	ctt Leu	ato Ile	ago Ser	gca Ala 360	Thr	. GJÀ : aaa	cgc Arg	gag Glu	rato Met 365	Ala	aat Asr	ata 1 Ile	ggt Gly	cag Gln 370	16280
					Pro					Pro					acc Thr	16328
				Thr					Tyr					/ Asr	ttg Leu	16376
			e Glr) Ala					1 Тут		aca Thr	16424
															c teu	16472

- 32 -

420	425		430	
acc gaa gat ccg Thr Glu Asp Pro 435	gcg caa gta gat Ala Gln Val Asq 440	t gct tta ttt p Ala Leu Phe 445	gat gca ggc gga c Asp Ala Gly Gly F	cat 16520 His 150
cag aac acg ttg Gln Asn Thr Leu	ata tca gga caa Ile Ser Gly Glr 455	a aac ctg aac n Asn Leu Asn 460	tgg aat aca cgc g Trp Asn Thr Arg C 465	ggt 16568 Gly
			aag ggc gcc aat g Lys Gly Ala Asn A 480	
		r Ser Asp Gly	aga agg ata tta a Arg Arg Ile Leu I 495	
			aca cag aga ata a Thr Gln Arg Ile 5	
			aac agc acg atc a Asn Ser Thr Ile ?	
			gcg ggt cgg gca (Ala Gly Arg Ala (545	
	His Trp Asp Se		gaa gat atc gac a Glu Asp Ile Asp 2 560	
		sp Asn Leu Ile	ggt tcc agt caa Gly Ser Ser Gln : 575	
gaa tta gac agc Glu Leu Asp Ser 580	aaa gga gaa at Lys Gly Glu Il 585	t att agt gag e Ile Ser Glu	gaa gag tac tat Glu Glu Tyr Tyr 590	ccc 16952 Pro
			cgg aca gaa gcc Arg Thr Glu Ala	
			gat gcc acc gga Asp Ala Thr Gly 625	
tat tat tac ggt Tyr Tyr Tyr Gly 630	Tyr Arg Tyr Ty	at cag cct tgg /r Gln Pro Trp 635	gta gga cga tgg Val Gly Arg Trp 640	tta 17096 Leu
agt gcc gat ccc Ser Ala Asp Pro 645	g gca gga aca gt o Ala Gly Thr Va 65	al Asp Gly Leu	aat tta tat cgg Asn Leu Tyr Arg 655	atg 17144 Met
			gat gga tta atg Asp Gly Leu Met 670	
			aat aaa gta gca Asn Lys Val Ala	

											gcg Ala					17288
											gca Ala					17336
											ata Ile					17384
											gta Val 750					17432
											agc Ser					17480
											act Thr					17528
											ttc Phe					17576
											agg Arg				agt Ser	17624
											cct Pro 830				gtc Val	17672
											tta Leu				tca Ser 850	17720
act Thr	tac Tyr	ata Ile	aaa Lys	tat Tyr 855	acc Thr	aag Lys	gac Asp	aaa Lys	tct Ser 860	aca Thr	gta Val	tgg Trp	gtc Val	tct Ser 865	act Thr	17768
				Glu										Leu	cat His	17816
gaa Glu	att Ile	aat Asn 885	atg Met	gat Asp	ctt Leu	tat Tyr	gag Glu 890	Phe	acc Thr	att Ile	gac Asp	gga Gly 895	caa Gln	aag Lys	cta Leu	17864
		Leu					Ser					Pro			tta Leu	17912
	Asp			_		Glu		_			Ile	_			cat His 930	17960
					Ala					Leu					ctt Leu	18008

PCT/EP99/01015 WO 99/42589

18055

aaa aat gta aaa cct tat aag aga taa cgaaaaatta atattettta Lys Asn Val Lys Pro Tyr Lys Arg 955 tctactttta atagccctct tgaacttaca ctcaaggggg ggaaaccaaa taagaaacca 18115 totttaataa caagocatga aagaatattt atttoatggo ttgattaott ttaacattoa 18175 atattaaata attaaaacaa tatctaacca attaaaataa caatacctta tttatcatat 18235 taaaatatca aatcagaaat taatgaattt aagggttett tatatttatt tetgagagca 18295 taggcacaat accttaccga tggcgctgga cgtgattcaa aatccagaaa tgctatattt 18355 tcatcaatat gggcagaata gcgcatttca ttgggagtca ttaaacttat cgcgacaccc 18415 gettttacca gatecaatet attagtaaaa teagggaeeg teaataaege taaattttgg 18475 tattcaggga gataattcaa tggcataaaa ttattgcatt gttttaaaaa agcactatta 18535 tgctgaacaa aaggaaaact agatattatt tcatcagcgt gactttctgg ttctaaaata 18595 tcatgggata cagcaagaga cagcatttga taagcaccat ctatcctgat gatatcatca 18655 ttatctggat aacattcagt cgtcacataa actgttatat cccctttcat taaggaagaa 18715 aataccgcat cttgccttat taaatcatca attagaaaat tgttgattat acaaatatcg 18775 cgataatgat aacgttgcac cgctcttttt acgaccgtag atattttatt aacatattct 18835 ccacttgtgc caataaccag tttgtctctg tttgataatt tataatttct acgacaattc 18895 caattattet caacttteag gateetttea taacaeggea geaactettg atatagtgee 18955 tttccctctt ctgtgagctt ggtttttccc ggtagtcgct caaacaattg acaccccaca 19015 cgctgttcca gttgatatac gagcctgcta agtggagaag gggtaataca aagcgtatcc 19075 gccgctaacg tgaatgactc tttcttagct gattccataa aatactttag ttgctttgaa 19135 caaaatatca tcacataccc tcttgttttc attccagaaa tagaatatta accatagaac 19195 atgacaacga tgtttctact ttgcattctt ttacattagg acatgcgtta atggacattg 19255 tataacttag ttaattattt cttgatattg atcatggtaa gttttcctca atacctacag 19375 aagtagatat tattttatet teeagtaate tategtttgg egaeggaggt egattettee 19435 attgggatat tcaacccatt cgccgccttt cttattaatt acagtgattt ttggcatttt 19495 ggtttcatcc aacttaggtt tataggtgat tttccattta gcacccggtg ttaacttcaa 19555 cctaaaggga tacataccaa cttcaccttg taagaatatt ctgtttggtc taccttcaac 19615 gactttcaaa atggggtaaa taaccgggct aaaatcaatc gtatccaatg catcaatttc 19675 gctgatattt gtccgggctg catcattgat aaatgcgatt aaatcggttg ctgaatacgg 19735 aatagcatet tteaetagat gaeggaeate ggtataaete aetgaeacaa aggeteggte 19795 aatetteeae ttacategae egecaecatt aaaaggtagt tttgeetgaa agtaaeeggt 19855 tttcggatca gcttttacat ccagacgtaa tccgttataa gttggtacct taaaaggcga 19915

catattggaa t	ctaaacgat	atttaaggca	atcttttgag	atatacacag	cggatacatg	19975
cggctgtgtg t	atttaggtg	cgactccttc	tacagtaatc	cactgattct	ctttgggagg	20035
agagagegge t	catttgggt	cagcacagcc	tgatattaaa	atcacggata	agacagataa	20095
gtatttcttg a	tatttatca	tggtaagttt	tecteaacte	ctacagcgtt	atctgcatgt	20155
gtgtccaatt c	cagatette	ctgtttatct	atttagaaat	aaataagcta	cgctgatagc	20215
attacttcat a	tttccatac	atgaatcgaa	aatcgacttc	ttgagtgccg	ttatcaattt	20275
tgccgcccgg a	tattcaacc	cactegeege	ctttcttatt	agtcaccgtg	accttcgcca	20335
ttttggtttc a	tccagctta	ggcttaaaaa	taattttcca	tttagctcct	ggagttaacg	20395
tgagttgaaa a	aggacgcatt	tttaatactt	caccttgtaa	gaatattctg	ttcgggcgac	20455
cttcaacgac t	ttcaaaaca	gggtaaataa	ccgggctaaa	atcaatcgta	ttcaatgtcg	20515
agattttgct a	atattcatc	tggactatgc	cattgataga	tgcgattaaa	ccggttgctg	20575
aatacggaat a	agcatctttc	accagatggc	tgacatcagt	ataactcacc	gatacaaagg	20635
cccggttaat t	ttccattta	catcgtcccc	ctccattaaa	aggtagtttt	gcttgaaaat	20695
aaccggtttg t	ggatcggcc	ttcactttca	gacgaagccc	attataggtc	ggcactttaa	20755
aaggcgacat a	attggaatcc	agacgatact	caaggcaatc	ctttgatatg	tattctgcgg	20815
atacatgtgg t	tcggtatat	ttcggcgcta	ccccttctac	cgtgatccat	tgattttctt	20875
taggagggga a	agcggctca	tttgggtcag	cacagcctga	tattaaaatc	actgacaaga	20935
caaataagta t	tttttaaca	tttatcatgg	taagttttcc	tcaattccta	cagcattatc	20995
cgcataaata t	cctgtcaag	aatagcgttc	attgatttcg	tcaccaaaga	aacaagatag	21055
taaaaatcct a	attaccacag	ataaaaaaca	ccgcttatgc	cgtgagtaat	agtgagttga	21115
g cgaca ggga t	tacagcagtg	catccccatc	aattagteee	tttgaataaa	gggaacagaa	21175
tttgaaattt o	ccgtcatacc	gtccatatta	cggaacttag	attatgatta	ttaaatcacc	21235
accaaatggc a	aagaaaaatt	ttcatttttt	aatttacgaa	gaatgaattt	gtaagaaagt	21295
gttacaaact t	taatagaaat	taatttactg	ttaatctaat	gaaggatgaa	attataaaaa	21355
taacccattt c	ctcagggaca	acaatccaca	atatatagaa	ccactggtcc	tcacttaatt	21415
teetgteagg a	agtagaaata	tcctgatgac	tcagtcgatg	acatacagca	atgtcattgg	21475
tattgagact a	accgactgtt	taataaattt	cttttgtctt	taatggcgag	atacaagtga	21535
ttcactattt a	agcactatc	gataaataag	attccaaaat	agcgccatat	cttacaccac	21595
tcataattct a	atgtataaca	attggttaaa	taggatcatg	tgtaacagga	ttatgaaacg	21655
ttatttatat o	caaatctatc	aattatttta	tatatagttt	cacagtcaca	ctcgctatct	21715
ggtaccttca t	taaccaactg	ccctccctgc	gctaccttct	gataacaaca	gctacactaa	21775
ctatacccgc g	gcctataatt	atgaccgtgt	gaaaattcag	cgtagttcac	cggccacgca	21835

aaataactac	acgaaatatt	gctccccaga	gaaacaccgt	tcgaggttgt	ttcaatgaaa	21895
catcaaggta	gagacaccta	tgtattatta	caagatatta	aaccctctgc	gattactcat	21955
aggaatgtac	gtaatactta	tacaggcaac	ttcacgtcat	ccagagaaaa	ttaagttgta	22015
caaaatagac	atcaactaat	atagtaatag	aaaatcccct	gaaaatagat	tcaggggatt	22075
taataaatta	accaaaaatc	ataataaaaa	tttatttcat	tattttagga	taaatattta	22135
attagcctaa	taatgaatta	ttacttaaag	taattcctaa	acaatcaaat	cggaaattaa	22195
taaattcaat	ggttcttgat	atttatgcct	gagagtataa	gcacaatatt	tcactgaggg	22255
tgtcggatgc	gatttaaaat	tcaaaaaggt	aatgccttta	tgaaggtcag	cagaacagag	22315
cacttcattc	ggtgtcatta	aacttatcgc	gatacctgat	ttcactaagt	ccaaccgatt	22375
agtaaaatca	ggcacagtta	gcaaagttaa	attctggtat	tcaggtagac	aattcagtga	22435
aataaaattg	ccgcactgtt	taaagaaagc	actattatgc	tgagccaaag	gcaatgtata	22495
tataatatta	tcagcgttac	tttccgatcc	taaaatatca	ttagctacag	caaggcgcaa	22555
agtctgataa	gttccatcca	ctctgataat	atcatcgttg	tcaggataat	gttcagtcgt	22615
cacatagacg	gttatatctt	ccttcattaa	agaagaaaat	attgtctctt	ttcttgccga	22675
atcatcaacc	agaaaattgt	tttttatgta	aatatcacga	taatgataac	gttgtaccgc	22735
tctttttatc	actgccgaaa	ttttattaat	atattctcca	cttgtcccga	tgaccagttt	22795
gccggtgttt	gctaattttc	cgcttctacg	ataatgccaa	ttatcctcaa	ctcgctgaat	22855
tctctcataa	cacggcaata	gctcctgata	taatgeette	ccctcttcag	tgagtttggt	22915
ccttcccggt	agtcgttcaa	atagttgaca	gcccacacgt	tgttccagtt	gatatatgat	22975
cctacttaga	ggagagggag	taatacaaag	cgtatccgcg	gctaaagtaa	atgactcttt	23035
tttcgctgac	tccataaaat	atttcaagct	ctttgaacaa	aatagcatca	tatatccttc	23095
ttattttaat	tcattgttcc	atccgaaata	gaatggaatg	ttaacaagaa	aacattacaa	23155
ctacttttct	tetttgcatt	atttaacatc	: aaagtatgca	ı ttaactgaga	ttgagtttta	23215
tcatctttat	. tcttaacagt	. tatcaaacaa	ttttcattat	tattgcaaaa	taaatacaac	23275
cccttcttat	gttacaataa	ı tgattataaa	gaaatttcad	atattatcat	taagtaataa	23335
tgggcacaat	: taaccattta	attaaacatt	tcaattggtt	gacaaagact	cattatgttc	23395
aacatgtaat	: gagcgcaatt	ttaacattaa	ataaattaca	a tagttcatat	tcattatcac	23455
tgagatcago	: tttt ttc gta	a tagtacatca	tgtgaacaat	accgtgccat	ttcctgccaa	23515
atcttattaa	aaagtcagtt	gcaaattttg	g catctgctt	ttttgcaaca	gctatttaaa	23575
gaaaacagtg	, agat agt gat	tatccgagag	g atcaagata	gtetgetett	tacgcacaaa	23635
ctgcaaacca	tttctatgca	a tatctcagct	atttctcaaa	a acctgtattt	aatcatctct	23695
tattccgato	gaacggaato	c attctctgat	tgattcatga	a tgtaaagaca	atatggatgt	23755
ttcatttact	tt atg att	tta aaa g	ga ata aat a	atg aat tcg	cct gta aaa	23806

Met	Ile	Leu	Lys	Gly	Ile	Asn	Met	Asn	Ser	Pro	Val	Lys
				960					965			-

gag ata cct gat gta tta aaa atc cag tgt ggt ttt cag tgt ctg aca Glu Ile Pro Asp Val Leu Lys Ile Gln Cys Gly Phe Gln Cys Leu Thr 970 975 980	23854
gat att age cac age tet tit aac gaa tit cac cag caa gia tee gaa Asp Ile Ser His Ser Ser Phe Asn Glu Phe His Gln Gln Val Ser Glu 985 990 995 1000	23902
cac ctc tcc tgg tcc gaa gca cac gac tta tat cat gat gca caa cag His Leu Ser Trp Ser Glu Ala His Asp Leu Tyr His Asp Ala Gln Gln 1005 1010 1015	23950
gcc caa aag gat aat cgg ctg tat gaa gcg cgt att ctt aaa cgc acg Ala Gln Lys Asp Asn Arg Leu Tyr Glu Ala Arg Ile Leu Lys Arg Thr 1020 1025 1030	23998
aat oot caa tta caa aat got gta cat ott god ato gta gog oot aat Asn Pro Gln Leu Gln Asn Ala Val His Leu Ala Ile Val Ala Pro Asn 1035 1040 1045	24046
gct gaa ctg ata ggc tat aac aac caa ttt agc ggc agg gcc agt caa Ala Glu Leu Ile Gly Tyr Asn Asn Gln Phe Ser Gly Arg Ala Ser Gln 1050 1055 1060	24094
tat gtc gcg ccg ggt acc gtt tcc tcc atg ttc tcc ccc gcc gct tat Tyr Val Ala Pro Gly Thr Val Ser Ser Met Phe Ser Pro Ala Ala Tyr 1065 1070 1075 1080	24142
ttg act gag ctt tat cgt gaa gca cgc aat tta cac gcc agc gat tcc Leu Thr Glu Leu Tyr Arg Glu Ala Arg Asn Leu His Ala Ser Asp Ser 1085 1090 1095	24190
gtt tat cgc ctg gat act cgc cgc cca gat ctc aaa tca atg gcg ctc Val Tyr Arg Leu Asp Thr Arg Arg Pro Asp Leu Lys Ser Met Ala Leu 1100 1105 1110	24238
agt caa caa aat atg gat acg gaa ctt tcc act ctc tct tta tcc aat Ser Gln Gln Asn Met Asp Thr Glu Leu Ser Thr Leu Ser Leu Ser Asn 1115 1120 1125	24286
gag cta tta ttg gaa agc att aaa act gag tct aag ctg gat aat tat Glu Leu Leu Glu Ser Ile Lys Thr Glu Ser Lys Leu Asp Asn Tyr 1130 1135 1140	24334
act caa gtg atg gaa atg ctc tcc gct ttc cgt cct tcc ggc gcg acg Thr Gln Val Met Glu Met Leu Ser Ala Phe Arg Pro Ser Gly Ala Thr 1145 1150 1155 1160	24382
cct tat cac gat gct tac gaa aat gtg cgt aaa gtt atc cag cta caa Pro Tyr His Asp Ala Tyr Glu Asn Val Arg Lys Val Ile Gln Leu Gln 1165 1170 1175	24430
gat cet ggg ett gag eaa tta aat get tea eea gee att gee ggg etg Asp Pro Gly Leu Glu Gln Leu Asn Ala Ser Pro Ala Ile Ala Gly Leu 1180 1185 1190	24478
atg cat caa gct tcc cta tta ggt att aac gct tca atc tca cct gag Met His Gln Ala Ser Leu Leu Gly Ile Asn Ala Ser Ile Ser Pro Glu 1195 1200 1205	24526
ttg ttt aat att ctg acg gag gag att act gaa ggt aat gct gag gaa Leu Phe Asn Ile Leu Thr Glu Glu Ile Thr Glu Gly Asn Ala Glu Glu	24574

- 38 -

1210	1215		1220	
ctt tat aag aaa Leu Tyr Lys Lys 1225	aat ttt ggt a Asn Phe Gly A 1230	at atc gaa ccg Asn Ile Glu Pro 1235	gct tca ctg gct at Ala Ser Leu Ala Me 124	t
Pro Glu Tyr Leu	aga cgt tat t Arg Arg Tyr T .245	tac aat tta agt Tyr Asn Leu Ser 1250	gat gaa gaa ctc ag Asp Glu Glu Leu Se 1255	c 24670 r
			caa gaa tat agt aa Gln Glu Tyr Ser As 1270	
aac caa ctc att Asn Gln Leu Ile 1275	Thr Pro Ile V	gtc aac agc aat Val Asn Ser Asn 280	gat ggc aca gtc aa Asp Gly Thr Val Ly 1285	g 24766 s
gta tat cga att Val Tyr Arg Ile 1290	acc cgc gaa t Thr Arg Glu T 1295	tat aca aca aat Iyr Thr Thr Asr	gcc aat caa gta ga Ala Asn Gln Val As 1300	c 24814 p
gtg gag ctg ttt Val Glu Leu Phe 1305	ccc tac ggt o Pro Tyr Gly 0 1310	gga gaa aat tat Gly Glu Asn Tyr 1315	cag tta aat tac aa Gln Leu Asn Tyr Ly 132	S
Phe Lys Asp Ser			a toc ato aaa tta aa 1 Ser Ile Iys Leu As 1335	
			g cet cag gte aae at a Pro Gln Val Asn Il 1350	
	His Ile Thr 1		t gat atc agt caa co r Asp Ile Ser Gln Pr 1365	
			t agt tct tgg gca ta r Ser Ser Trp Ala Ty 1380	
			c caa tac tct ttc ct n Gln Tyr Ser Phe Le 5 140	eu
	_		t gcg aca gaa tta to g Ala Thr Glu Leu Se 1415	
	Glu Ser Ile		t aat cag caa ctg ga l Asn Gln Gln Leu As 1430	
	Val Leu Gly		g act aaa tat tat a u Thr Lys Tyr Tyr Me 1445	
			a ata cta tgc aat go u Ile Leu Cys Asn A 1460	
			et agc caa ttt gat co To Ser Gln Phe Asp A 5 14	rg

ctg ttt aat acg cca tta ctg aac ggc caa tat ttt tct acc gga gat Leu Phe Asn Thr Pro Leu Leu Asn Gly Gln Tyr Phe Ser Thr Gly Asp 1485 1490 1495	25390
gaa gag att gat tta aat cca ggt agt act ggc gat tgg cgt aaa tcc Glu Glu Ile Asp Leu Asn Pro Gly Ser Thr Gly Asp Trp Arg Lys Ser 1500 1505 1510	25438
gtg ctt aaa cgt gca ttt aat atc gat gat att tcc ctc tac cgc ctg Val Leu Lys Arg Ala Phe Asn Ile Asp Asp Ile Ser Leu Tyr Arg Leu 1515 1520 1525	25486
ctt aaa att acc aac cat aat aat caa gat gga aag att aaa aat aac Leu Lys Ile Thr Asn His Asn Asn Gln Asp Gly Lys Ile Lys Asn Asn 1530 1535 1540	25534
tta aat aat ctt tct gat tta tat att ggg aaa tta ctg gca gaa att Leu Asn Asn Leu Ser Asp Leu Tyr Ile Gly Lys Leu Leu Ala Glu Ile 1545 1550 1560	25582
cat caa tta acc att gat gaa ttg gat tta ttg ctg gtt gcc gtg ggt His Gln Leu Thr Ile Asp Glu Leu Asp Leu Leu Leu Val Ala Val Gly 1565 1570 1575	25630
gaa gga gaa act aat tta too got ato agt gat aaa caa otg gog goa Glu Gly Glu Thr Asn Leu Ser Ala Ile Ser Asp Lys Gln Leu Ala Ala 1580 1585 1590	25678
ctg atc aga aaa ctc aat acc att acc gtc tgg cta cag aca cag aag Leu Ile Arg Lys Leu Asn Thr Ile Thr Val Trp Leu Gln Thr Gln Lys 1595 1600 1605	25726
tgg agt gcg ttc caa tta ttt gtt atg act tcc acc agc tat aac aaa Trp Ser Ala Phe Gln Leu Phe Val Met Thr Ser Thr Ser Tyr Asn Lys 1610 1615 1620	25774
acg ctg acg cct gaa att aag aat ctg ctg gat acc gtc tac cac ggt Thr Leu Thr Pro Glu Ile Lys Asn Leu Leu Asp Thr Val Tyr His Gly 1625 1630 1635 1640	25822
tta caa ggc ttt gat aaa gac aag gca aat tta ctg cat gtt atg gcg Leu Gln Gly Phe Asp Lys Asp Lys Ala Asn Leu Leu His Val Met Ala 1645 1650 1655	25870
ccc tat att gcg gcc acc tta caa tta tca tcg gaa aat gtc gcc cat Pro Tyr Ile Ala Ala Thr Leu Gln Leu Ser Ser Glu Asn Val Ala His 1660 1665 1670	25918
tct gtg ctg ctt tgg gca gac aag tta aag ccc ggc gac ggc gca atg Ser Val Leu Leu Trp Ala Asp Lys Leu Lys Pro Gly Asp Gly Ala Met 1675 1680 1685	25966
aca gcc gaa aaa ttc tgg gac tgg ttg aat act caa tat acg cca gat Thr Ala Glu Lys Phe Trp Asp Trp Leu Asn Thr Gln Tyr Thr Pro Asp 1690 1695 1700	26014
tca tcg gaa gta tta gca aca cag gaa cat att gtt cag tat tgt cag Ser Ser Glu Val Leu Ala Thr Gln Glu His Ile Val Gln Tyr Cys Gln 1705 1710 1715 1720	26062
gcg ttg gcg caa tta gaa atg gtt tac cat tcc acc ggt atc aat gaa Ala Leu Ala Gln Leu Glu Met Val Tyr His Ser Thr Gly Ile Asn Glu	26110

WO 99/42589 PCT/EP99/01015

aac gcc ttc cgc ctg ttt gtg aca aaa cca gag atg ttt ggc tcg tca Asn Ala Phe Arg Leu Phe Val Thr Lys Pro Glu Met Phe Gly Ser Ser 1740 1745 1750	26158
act gag gca gta cet gcg cat gat gca ett tea etg ate atg etg acg Thr Glu Ala Val Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr 1755 1760 1765	26206
cgt ttt gca gat tgg gtt aat gcg tta ggc gaa aaa gcc tct tcc gta Arg Phe Ala Asp Trp Val Asn Ala Leu Gly Glu Lys Ala Ser Ser Val 1770 1775 1780	26254
cta gcg gca ttt gaa gct aac agt tta acg gca gaa caa ttg gct gat Leu Ala Ala Phe Glu Ala Asn Ser Leu Thr Ala Glu Gln Leu Ala Asp 1785 1790 1795 1800	26302
gcc atg aat ctt gat gct aat ttg cta ttg caa gcc agt act caa gca Ala Met Asn Leu Asp Ala Asn Leu Leu Gln Ala Ser Thr Gln Ala 1805 1810 1815	26350
caa aac cat caa cat ctt ccc cca gtg acg caa aaa aat gct ttc tcc Gln Asn His Gln His Leu Pro Pro Val Thr Gln Lys Asn Ala Phe Ser 1820 1825 1830	26398
tgt tgg aca tct atc gac act atc ctg caa tgg gtt aat gtt gca caa Cys Trp Thr Ser Ile Asp Thr Ile Leu Gln Trp Val Asn Val Ala Gln 1835 1840 1845	26446
caa ttg aat gtc gcc cca cag gga gtt tcc gct ttg gtc ggg ctg gat Gln Leu Asn Val Ala Pro Gln Gly Val Ser Ala Leu Val Gly Leu Asp 1850 1855 1860	26494
tat att caa tta aat caa aaa atc ccc acc tat gcc cag tgg gaa agt Tyr Ile Gln Leu Asn Gln Lys Ile Pro Thr Tyr Ala Gln Trp Glu Ser 1865 1870 1875 1880	26542
gct ggg gaa ata ttg act gcc gga ttg aat tca caa cag gct gat ata Ala Gly Glu Ile Leu Thr Ala Gly Leu Asn Ser Gln Gln Ala Asp Ile 1885 1890 1895	26590
tta cac gct ttt ttg gac gaa tct cgc agt gcc gca tta agc acc tac Leu His Ala Phe Leu Asp Glu Ser Arg Ser Ala Ala Leu Ser Thr Tyr 1900 1905 1910	26638
tat atc cgt caa gtc gcc aag cca gcg gca gcc ata aaa agc cgt gat Tyr Ile Arg Gln Val Ala Lys Pro Ala Ala Ala Ile Lys Ser Arg Asp 1915 1920 1925	26686
gac ttg tac caa tac tta cta att gat aat cag gtt tcc gct gca atc Asp Leu Tyr Gln Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Ala Ile 1930 1935 1940	26734
aaa act acc cgg att gcc gaa gcc att gcc agc att caa ctg tac gtc Lys Thr Thr Arg Ile Ala Glu Ala Ile Ala Ser Ile Gln Leu Tyr Val 1945 1950 1955 1960	26782
aac cgc acg ctg gaa aat gta gaa gaa aat gcc cat tca ggg gtt atc Asn Arg Thr Leu Glu Asn Val Glu Glu Asn Ala His Ser Gly Val Ile 1965 1970 1975	26830
agc cgt cag ttc ttt atc gac tgg gac aaa tat aac aaa cgc tac agc Ser Arg Gln Phe Phe Ile Asp Trp Asp Lys Tyr Asn Lys Arg Tyr Ser 1980 1985 1990	26878
acc tgg gcg ggt gtt tct caa tta gtt tac tac ccg gaa aac tat att	26926

Thr Trp Ala Gly Val Ser Gln Leu Val Tyr Tyr Pro Glu Asn Tyr Ile 1995 2000 2005	
gat ccc acc atg cgt atc gga caa acc aaa atg atg gac gca tta ttg 20 Asp Pro Thr Met Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu 2010 2015 2020	6974
caa tcc gtc agc caa agc caa tta aat gcc gat act gtc gaa gac gcc 2'Gln Ser Val Ser Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala 2025 2030 2035 2040	7022
Phe Met Ser Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val 2045 2050 2055	70 7 0
att agc gcg tat cac gat aat att aac aac gat caa ggg ctg acc tat 2' Ile Ser Ala Tyr His Asp Asn Ile Asn Asp Gln Gly Leu Thr Tyr 2060 2065 2070	7118
ttt atc ggc ctc agt gaa act gat acc ggt gaa tac tat tgg cgc agt 2°. Phe Ile Gly Leu Ser Glu Thr Asp Thr Gly Glu Tyr Tyr Trp Arg Ser 2075 2080 2085	7166
gtc gat cac agt aaa ttc agc gac ggt aaa ttc gcc gct aat gcc tgg 2' Val Asp His Ser Lys Phe Ser Asp Gly Lys Phe Ala Ala Asn Ala Trp 2090 2095 2100	7214
agt gaa tgg cac aaa att gat tgt cca att aat cct tac cga agc act Ser Glu Trp His Lys Ile Asp Cys Pro Ile Asn Pro Tyr Arg Ser Thr 2105 2110 2115 2120	7262
atc cgt cct gtg atg tac aaa tcc cgc ttg tat ctg ctc tgg ttg gaa 2' Ile Arg Pro Val Met Tyr Lys Ser Arg Leu Tyr Leu Leu Trp Leu Glu 2125 2130 2135	7310
caa aag gag atc act aaa caa aca gga aat agc aaa gat ggc tat caa 27 Gln Lys Glu Ile Thr Lys Gln Thr Gly Asn Ser Lys Asp Gly Tyr Gln 2140 2145 2150	7358
acc gag aca gat tat cgt tat gag cta aaa ttg gcg cat atc cgt tat 2°. Thr Glu Thr Asp Tyr Arg Tyr Glu Leu Lys Leu Ala His Ile Arg Tyr 2155 2160 2165	7406
gac ggt acc tgg aat acg cca atc act ttt gat gtc aat gaa aaa ata 27 Asp Gly Thr Trp Asn Thr Pro Ile Thr Phe Asp Val Asn Glu Lys Ile 2170 2175 2180	7454
tcc aag cta gaa ctg gca aaa aat aaa gcg cct ggg ctc tat tgt gct 2' Ser Lys Leu Glu Leu Ala Lys Asn Lys Ala Pro Gly Leu Tyr Cys Ala 2185 2190 2195 2200	7502
ggt tat caa ggt gaa gat acg ttg ctg gtt atg ttt tat aac caa caa 27 Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Asn Gln Gln 2205 2210 2215	7550
gat aca ctc gat agt tat aaa acc gct tca atg caa ggg cta tat atc 2' Asp Thr Leu Asp Ser Tyr Lys Thr Ala Ser Met Gln Gly Leu Tyr Ile 2220 2225 2230	7598
ttt gcc gat atg gaa tat aaa gat atg acc gat gga caa tac aaa tct 2' Phe Ala Asp Met Glu Tyr Lys Asp Met Thr Asp Gly Gln Tyr Lys Ser 2235 2240 2245	7646
tat cgg gac aac agc tat aaa caa ttc gat act aat agt gtc aga aga 2° Tyr Arg Asp Asn Ser Tyr Lys Gln Phe Asp Thr Asn Ser Val Arg Arg	7694

2250	2255	2260		
gtg aat aac cgc Val Asn Asn Arg 2265	tat gca gag gat Tyr Ala Glu Asp 2270	tat gaa att ccc Tyr Glu Ile Pro 2275	tca tcg gta aat Ser Ser Val Asn 2280	27742
Ser Arg Lys Gly	tat gat tgg gga Tyr Asp Trp Gly 2285	gat tat tat ctc Asp Tyr Tyr Leu 2290	agt atg gta tat Ser Met Val Tyr 2295	27790
aac gga gat att Asn Gly Asp Ile 2300	cca act att agt Pro Thr Ile Ser 2	tac aaa gcc aca Tyr Lys Ala Thr 305	tca agt gat tta Ser Ser Asp Leu 2310	27838
aaa atc tat atc Lys Ile Tyr Ile 2315	tcg cca aaa tta Ser Pro Lys Leu 2320	Arg Ile Ile His	aat gga tat gaa Asn Gly Tyr Glu 2325	27886
ggg cag caa cgc Gly Gln Gln Arg 2330	aat caa tgc aat Asn Gln Cys Asn 2335	cta atg aat aaa Leu Met Asn Lys 2340	tat ggc aaa cta Tyr Gly Lys Leu	27934
ggt gat aaa ttt Gly Asp Lys Phe 2345	att gtt tat act Ile Val Tyr Thr 2350	agc ttg gga gtt Ser Leu Gly Val 2355	aat cca aat aat Asn Pro Asn Asn 2360	27982
Ser Ser Asn Lys	ctg atg ttt tac Leu Met Phe Tyr 2365	ccc gtt tat caa Pro Val Tyr Gln 2370	tat aac gga aat Tyr Asn Gly Asn 2375	28030
gtc agt ggg ctt Val Ser Gly Leu 2380	agt caa ggg aga Ser Gln Gly Arg 2	tta cta ttc cac Leu Leu Phe His 2385	cgt gac acc aat Arg Asp Thr Asn 2390	28078
tat tca tct aaa Tyr Ser Ser Lys 2395	gta gaa gct tgg Val Glu Ala Trp 2400	Ile Pro Gly Ala	gga cgt tct cta Gly Arg Ser Leu 2405	28126
acc aat ccg aat Thr Asn Pro Asn 2410	gct gcc att ggt Ala Ala Ile Gly 2415	gat gat tat gct Asp Asp Tyr Ala 2420	aca gac tcg tta Thr Asp Ser Leu	28174
aac aaa ccg aat Asn Lys Pro Asn 2425	gat ctt aag caa Asp Leu Lys Gln 2430	tac gtc tat atg Tyr Val Tyr Met 2435	act gac agt aaa Thr Asp Ser Lys 2440	28222
Gly Thr Ala Thr	gat gtc tca gga Asp Val Ser Gly 2445	cca gta gat atc Pro Val Asp Ile 2450	aat act gca att Asn Thr Ala Ile 2455	28270
tcc ccg gca aaa Ser Pro Ala Lys 2460	gtt cag gta aca Val Gln Val Thr 2	gta aaa gcc ggt Val Lys Ala Gly 2465	agc aaa gaa caa Ser Lys Glu Gln 2470	28318
acg ttt acc gcg Thr Phe Thr Ala 2475	gat aaa aat gtc Asp Lys Asn Val 2480	Ser Ile Gln Pro	tcc cct agc ttt Ser Pro Ser Phe 2485	28366
gat gaa atg aat Asp Glu Met Asn 2490	tat caa ttt aat Tyr Gln Phe Asn 2495	gct ctc gaa ata Ala Leu Glu Ile 2500	gat ggc tca agt Asp Gly Ser Ser	28414
ctg aat ttt act Leu Asn Phe Thr 2505	aac aat tca gcc Asn Asn Ser Ala 2510	agt att gat att Ser Ile Asp Ile 2515	acc ttt acc gca Thr Phe Thr Ala 2520	28462

ttt gca gag gat gga cgt aaa ctg ggt tat gaa agt ttc agt att cct Phe Ala Glu Asp Gly Arg Lys Leu Gly Tyr Glu Ser Phe Ser Ile Pro 2525 2530 2535	28510
att acc cgc aag gtg agt act gat aat tcc ctg acc ctg cgc cat aat Ile Thr Arg Lys Val Ser Thr Asp Asn Ser Leu Thr Leu Arg His Asn 2540 2545 2550	28558
gaa aat ggt gcg caa tat atg caa tgg gga gtc tat cgc att cgt ctt Glu Asn Gly Ala Gln Tyr Met Gln Trp Gly Val Tyr Arg Ile Arg Leu 2555 2560 2565	28606
aat act tta ttt gct cgc caa tta gtt gcg cga gcc act acc ggt att Asn Thr Leu Phe Ala Arg Gln Leu Val Ala Arg Ala Thr Thr Gly Ile 2570 2575 2580	28654
gat acg att ctg agt atg gaa act cag aat att cag gaa cca cag tta Asp Thr Ile Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu 2585 2590 2595 2600	28702
ggc aaa ggt ttc tac gct acg ttc gtg ata cct ccg tat aac cca tca Gly Lys Gly Phe Tyr Ala Thr Phe Val Ile Pro Pro Tyr Asn Pro Ser 2605 2610 2615	28750
act cat ggt gat gaa cgt tgg ttt aag ctt tat atc aaa cat gtt gtt Thr His Gly Asp Glu Arg Trp Phe Lys Leu Tyr Ile Lys His Val Val 2620 2625 2630	28798
gat aat aat toa oat att ato tat toa ggt oag ota aaa gat aca aat Asp Asn Asn Ser His Ile Ile Tyr Ser Gly Gln Leu Lys Asp Thr Asn 2635 2640 2645	28846
ata agc acc acg tta ttt atc cct ctt gat gat gtt cca ttg aac caa Ile Ser Thr Thr Leu Phe Ile Pro Leu Asp Asp Val Pro Leu Asn Gln 2650 2655 2660	28894
gat tac age gcc aag gtt tac atg acc ttc aag aaa tca cca tca gat Asp Tyr Ser Ala Lys Val Tyr Met Thr Phe Lys Lys Ser Pro Ser Asp 2665 2670 2675 2680	28942
ggt acc tgg tgg ggc cct cac ttt gtt aga gat gat aaa gga ata gta Gly Thr Trp Trp Gly Pro His Phe Val Arg Asp Asp Lys Gly Ile Val 2685 2690 2695	28990
aca ata aac cct aaa tcc att ttg acc cac ttt gag agc gtc aat gtc Thr Ile Asn Pro Lys Ser Ile Leu Thr His Phe Glu Ser Val Asn Val 2700 2705 2710	29038
ctg aat aat att agt agc gaa cca atg gat ttc agc ggc gct aac agc Leu Asn Asn Ile Ser Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ser 2715 2720 2725	29086
ctc tat ttt tgg gaa ctg ttc tac tat acc ccg atg ctg gtt gcc caa Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro Met Leu Val Ala Gln 2730 2735 2740	29134
cgt ttg ttg cat gag caa aac ttt gat gaa gcg aac cgc tgg ctg aaa Arg Leu Leu His Glu Gln Asn Phe Asp Glu Ala Asn Arg Trp Leu Lys 2745 2750 2755 2760	29182
tat gtc tgg agc cca tcc ggg tat att gtt cac ggc cag att cag aat Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val His Gly Gln Ile Gln Asn 2765 2770 2775	29230

tat caa tgg aac gtc cgc ccg tta ttg gaa gat acc agt tgg aac agt Tyr Gln Trp Asn Val Arg Pro Leu Leu Glu Asp Thr Ser Trp Asn Ser 2780 2785 2790	29278
gat cct ttg gat tcc gtc gat cct gac gcg gta gcg cag cac gat ccg Asp Pro Leu Asp Ser Val Asp Pro Asp Ala Val Ala Gln His Asp Pro 2795 2800 2805	29326
atg cac tat aaa gtt tca acc ttt atg cgc acc ctt gat ctg ttg atc Met His Tyr Lys Val Ser Thr Phe Met Arg Thr Leu Asp Leu Leu Ile 2810 2815 2820	29374
gcg cgc ggc gac cat gct tac cgc caa ttg gag cgc gat acg ctt aac Ala Arg Gly Asp His Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Asn 2825 2830 2835 2840	29422
gaa gcg aag atg tgg tat atg caa gcg ctg cat ctg tta ggc gat aaa Glu Ala Lys Met Trp Tyr Met Gln Ala Leu His Leu Leu Gly Asp Lys 2845 2850 2855	29470
cct tat ctg ccg ctg agt acc aca tgg aat gat cca cga ctg gac aaa Pro Tyr Leu Pro Leu Ser Thr Thr Trp Asn Asp Pro Arg Leu Asp Lys 2860 2865 2870	29518
gcc gcg gat att act acc caa agt gct cat tcc agc tca ata gtc gct Ala Ala Asp Ile Thr Thr Gln Ser Ala His Ser Ser Ser Ile Val Ala 2875 2880 2885	29566
ttg cgg cag agt aca ccg gcg ctt tta tca ttg cgc agc gcc aat acc Leu Arg Gln Ser Thr Pro Ala Leu Leu Ser Leu Arg Ser Ala Asn Thr 2890 2895 2900	29614
ctg acc gat ctc ttc ctg ccg caa atc aat gaa gtg atg atg aat tac Leu Thr Asp Leu Phe Leu Pro Gln Ile Asn Glu Val Met Met Asn Tyr 2905 2910 2915 2920	29662
tgg caa aca tta gct cag aga gta tac aac ctg cgc cac aac ctc tct Trp Gln Thr Leu Ala Gln Arg Val Tyr Asn Leu Arg His Asn Leu Ser 2925 2930 2935	29710
atc gac ggt cag ccg tta tat ctg cca atc tat gcc aca ccg gcg gac Ile Asp Gly Gln Pro Leu Tyr Leu Pro Ile Tyr Ala Thr Pro Ala Asp 2940 2945 2950	29 75 8
ccg aaa gcg tta ctc agc gcc gct gtt gcc act tct caa ggt gga ggc Pro Lys Ala Leu Leu Ser Ala Ala Val Ala Thr Ser Gln Gly Gly Gly 2955 2960 2965	29806
aag ctg ccg gag tca ttt atg tcc ctg tgg cgt ttc ccg cac atg ctg Lys Leu Pro Glu Ser Phe Met Ser Leu Trp Arg Phe Pro His Met Leu 2970 2975 2980	29854
gaa aat gct cgc agc atg gtt agc cag ctc acc caa ttc ggc tcc acg Glu Asn Ala Arg Ser Met Val Ser Gln Leu Thr Gln Phe Gly Ser Thr 2985 2990 2995 3000	29902
tta caa aat att atc gaa cgt cag gac gca gaa gcg ctc aat gcg tta Leu Gln Asn Ile Ile Glu Arg Gln Asp Ala Glu Ala Leu Asn Ala Leu 3005 3010 3015	29950
tta caa aat cag gcc gca gag ctg ata ttg act aac ctg agt att caa Leu Gln Asn Gln Ala Ala Glu Leu Ile Leu Thr Asn Leu Ser Ile Gln 3020 3025 3030	29998
gac aaa acc att gaa gaa ctg gat gcc gag aaa acc gtg ctg gaa aaa	30046

Asp Lys Thr Ile 3035		Asp Ala Glu 1 040	Lys Thr Val Leu 3045	Glu Lys
tcc aaa gcg gga Ser Lys Ala Gly 3050	gca caa tcg c Ala Gln Ser A 3055	cgc ttt gat : Arg Phe Asp :	agc tat agc aaa Ser Tyr Ser Lys 3060	ctg cat 30094 Leu His
gat gaa aac atc Asp Glu Asn Ile 3065	aac gcc ggt g Asn Ala Gly (3070	Glu Asn Gln .	gct atg acg cta Ala Met Thr Leu 075	cga gcg 30142 Arg Ala 3080
Ser Ala Ala Gly			gca tcc cgt ctg Ala Ser Arg Leu 3	
	Leu Val Pro		ggc ttc gcc ggt Gly Phe Ala Gly 3110	
	Ala Ile Ala		ggc tat gta atg Gly Tyr Val Met 3125	
			aaa att agc caa Lys Ile Ser Gln 3140	
		Glu Trp Glu	att cag cgt aat Ile Gln Arg Asn 3155	
			ctt aaa tcg ctg Leu Lys Ser Leu	
	Ala Val Leu		agc ctg aaa acc Ser Leu Lys Thr 3190	
	Ala Gln Leu		caa cgt aag ttc Gln Arg Lys Phe 3205	
			ctg gca gca att Leu Ala Ala Ile 3220	
		Ala Arg Cys	tta atg gca gag Leu Met Ala Glu 3235	
			cgc ttt att aaa Arg Phe Ile Lys	
	y Thr Tyr Ala		gca ggt gaa acc Ala Gly Glu Thr 3270	Leu Met
	a Gln Met Glu		tta aga cgc gat Leu Arg Arg Asp 3285	
			gcc gaa att tat Ala Glu Ile Tyr	

3290	3295	3300		
tta ccg caa gat Leu Pro Gln Asp 3305	aaa ggc cca tto Lys Gly Pro Phe 3310	tcc ctg acg caa Ser Leu Thr Gln 3315	gaa atc gag aag Glu Ile Glu Lys 3320	30862
cig gtg aat gca Leu Val Asn Ala	a ggt tca ggc ago a Gly Ser Gly Ser 3325	gcc ggc agt ggt Ala Gly Ser Gly 3330	aat aat ttg Asn Asn Asn Leu 3335	30910
gca ttt ggc gcc Ala Phe Gly Ala 3340	a Gly Thr Asp Thi	aaa act tct ttg Lys Thr Ser Leu 3345	cag gca tcc att Gln Ala Ser Ile 3350	30958
tca tta gct ga Ser Leu Ala As 3355	t tta aaa att cg o Leu Lys Ile Arc 336	t gag gat tac ccg g Glu Asp Tyr Pro O	gaa tct att ggc Glu Ser Ile Gly 3365	31006
aaa atc cga cg Lys Ile Arg Ar 3370	c atc aaa cag atc g Ile Lys Gln Ile 3375	c agc gtt acc ctg e Ser Val Thr Leu 3380	ccg gcg cta ttg Pro Ala Leu Leu	31054
gga cct tat ca Gly Pro Tyr Gl 3385	g gat gtg cag gc n Asp Val Gln Al 3390	a ata tta tct tac a Ile Leu Ser Tyr 3395	ggc gat aaa gcc Gly Asp Lys Ala 3400	31102
gga tta gcg aa Gly Leu Ala As	c ggc tgt gca gc n Gly Cys Ala Al 3405	g ctg gcc gtt tcc a Leu Ala Val Ser 3410	cac ggt acg aat His Gly Thr Asn 3415	31150
gac agc ggt ca Asp Ser Gly Gl 342	n Phe Gln Leu As	it ttc aac gat ggc pp Phe Asn Asp Gly 3425	aaa ttc ctg ccg Lys Phe Leu Pro 3430	31198
ttt gaa ggt at Phe Glu Gly Il 3435	c gcc att gat ca e Ala Ile Asp Gl 344	aa ggt acg cta aca In Gly Thr Leu Thr 40	ctg agt ttt cct Leu Ser Phe Pro 3445	31246
aat gca tca ad Asn Ala Ser Th 3450	eg cca gcc aaa go nr Pro Ala Lys Gl 3455	gt aaa caa gcc act ly Lys Gln Ala Thr 3460	Met Leu Lys Thr	31294
ctg aac gat a Leu Asn Asp I 3465	ic att ttg cat at le Ile Leu His I 3470	tt cgc tac acc att le Arg Tyr Thr Ile 3475	aag taa E Lys	31336
ccatcccaac ac	agaactaa gacaggc	ccc gaatcggggt ctq	ggtaagga gtttct atg Met	
cag aat tca c Gln Asn Ser G 3480	ag aca ttc agc a ln Thr Phe Ser M 3485	tg acc gag ctg tca et Thr Glu Leu Se: 3490	a tta cct aag ggc r Leu Pro Lys Gly 3495	31443
ggc ggc gcc a Gly Gly Ala I	tt acc ggt atg g le Thr Gly Met G 3500	gt gaa gca tta ac ly Glu Ala Leu Th 3505	g ccg gcc ggg ccg r Pro Ala Gly Pro 3510	31491
Asp Gly Met A	ca gcc tta tcg c la Ala Leu Ser L 15	tg cca ttg ccc at eu Pro Leu Pro Il 3520	t tct gcc gga cgt e Ser Ala Gly Arg 3525	31539
ggt tat gcc c Gly Tyr Ala F 3530	ro Ser Leu Thr L	tg aac tac aac ag eu Asn Tyr Asn Se 335	c gga acc ggt aac r Gly Thr Gly Asn 3540	31587

agc ccg ttc ggt ctc ggt tgg gac tgt aac gtc atg aca att cgt cgt Ser Pro Phe Gly Leu Gly Trp Asp Cys Asn Val Met Thr Ile Arg Arg 3545 3550 3555	31635
cgc acc agt acc ggc gtg ccg aat tat gat gaa acc gat act ttt ctg Arg Thr Ser Thr Gly Val Pro Asn Tyr Asp Glu Thr Asp Thr Phe Leu 3560 3565 3570 3575	31683
ggg ccg gaa ggt gaa gtg ttg gtc gta gca tta aat gag gca ggt caa Gly Pro Glu Gly Glu Val Leu Val Val Ala Leu Asn Glu Ala Gly Gln 3580 3585 3590	31731
gct gat atc cgc agt gaa tcc tca tta cag ggc atc aat ttg ggg atg Ala Asp Ile Arg Ser Glu Ser Ser Leu Gln Gly Ile Asn Leu Gly Met 3595 3600 3605	31779
acc ttc acc gtt acc ggt tat cgc tcc cgt ttg gaa agc cac ttt agc Thr Phe Thr Val Thr Gly Tyr Arg Ser Arg Leu Glu Ser His Phe Ser 3610 3615 3620	31827
cgg ttg gaa tac tgg caa ccc caa aca aca ggc gca acc gat ttc tgg Arg Leu Glu Tyr Trp Gln Pro Gln Thr Thr Gly Ala Thr Asp Phe Trp 3625 3630 3635	31875
ctg ata tac agc ccc gac gga caa gcc cat tta ctg ggc aaa aat cct Leu Ile Tyr Ser Pro Asp Gly Gln Ala His Leu Leu Gly Lys Asn Pro 3640 3645 3650 3655	31923
caa gca cgc atc agc aat cca cta aat gtt aac caa aca gcg caa tgg Gln Ala Arg Ile Ser Asn Pro Leu Asn Val Asn Gln Thr Ala Gln Trp 3660 3665 3670	31971
cta ttg gaa gcc tcg gta tca tcc cac ggc gag cag att tat tat cag Leu Leu Glu Ala Ser Val Ser Ser His Gly Glu Gln Ile Tyr Tyr Gln 3675 3680 3685	32019
tat cga gcc gaa gat gaa act gat tgc gaa act gac gaa ctc aca gcc Tyr Arg Ala Glu Asp Glu Thr Asp Cys Glu Thr Asp Glu Leu Thr Ala 3690 3695 3700	32067
cac ccg aac aca acc gtc cag cgc tac ctg caa gta gta cat tac ggt His Pro Asn Thr Thr Val Gln Arg Tyr Leu Gln Val Val His Tyr Gly 3705 3710 3715	32115
aat cta acc gcc agc gaa gta ttt ccc acg cta aat gga gat gat cca Asn Leu Thr Ala Ser Glu Val Phe Pro Thr Leu Asn Gly Asp Asp Pro 3720 3725 3730 3735	32163
ctc aaa tct ggc tgg ttg ttc tgt tta gta ttt gat tac ggt gag cgc Leu Lys Ser Gly Trp Leu Phe Cys Leu Val Phe Asp Tyr Gly Glu Arg 3740 3745 3750	32211
aaa aac agc tta tct gaa atg ccg cca ttt aaa gcc aca agt aac tgg Lys Asn Ser Leu Ser Glu Met Pro Pro Phe Lys Ala Thr Ser Asn Trp 3755 3760 3765	32259
ctt tgc cgc aaa gac cgt ttt tcc cgt tat gaa tac ggt ttt gca ttg Leu Cys Arg Lys Asp Arg Phe Ser Arg Tyr Glu Tyr Gly Phe Ala Leu 3770 3775 3780	32307
cgc acc cgg cgc tta tgt cgc caa ata ctg atg ttt cac cgt ctg caa Arg Thr Arg Arg Leu Cys Arg Gln Ile Leu Met Phe His Arg Leu Gln 3785 3790 3795	32355
acc ctg tct ggt cag gca aaa ggc gac gat gaa ccc gca tta gtt tca	32403

Thr Leu Ser Gly Gln Ala Lys Gly Asp Asp Glu Pro Ala Leu Val Ser 3800 3805 3810 3815	
cgt ctg ata ctg gat tat gac gaa aac gcg gtg gtc agt acg ctc gtt Arg Leu Ile Leu Asp Tyr Asp Glu Asn Ala Val Val Ser Thr Leu Val 3820 3825 3830	32451
tet gte ege ega gtg gga eat gag eaa gat gge aca acg geg gte gee Ser Val Arg Arg Val Gly His Glu Gln Asp Gly Thr Thr Ala Val Ala 3835 3840 3845	32499
ctg ccg cca ttg gaa ctg gct tat cag cct ttt gaa cca gaa caa aaa Leu Pro Pro Leu Glu Leu Ala Tyr Gln Pro Phe Glu Pro Glu Gln Lys 3850 3855 3860	32547
gca ctc tgg cga cca atg gat gta ctg gcg aat ttc aac acc atc caa Ala Leu Trp Arg Pro Met Asp Val Leu Ala Asn Phe Asn Thr Ile Gln 3865 3870 3875	32595
cgc tgg caa ctg ctt gat ctg caa ggc gaa ggc gta ccc ggt att ctg Arg Trp Gln Leu Leu Asp Leu Gln Gly Glu Gly Val Pro Gly Ile Leu 3880 3885 3890 3895	32643
tat cag gat aaa aat ggc tgg tgg tat cga tct gct caa cgt cag aca Tyr Gln Asp Lys Asn Gly Trp Trp Tyr Arg Ser Ala Gln Arg Gln Thr 3900 3905 3910	32691
ggg gaa gag atg aat gcg gtc acc tgg ggc aaa atg caa ctc ctt cct Gly Glu Glu Met Asn Ala Val Thr Trp Gly Lys Met Gln Leu Leu Pro 3915 3920 3925	32739
atc acg ccc gct att cag gat aac gcc tca ctg atg gat att aat ggt Ile Thr Pro Ala Ile Gln Asp Asn Ala Ser Leu Met Asp Ile Asn Gly 3930 3935 3940	32 7 87
gat ggg caa ctg gat tgg gtt atc acc ggt ccg ggg cta agg ggt tat Asp Gly Gln Leu Asp Trp Val Ile Thr Gly Pro Gly Leu Arg Gly Tyr 3945 3950 3955	32835
cac agc cag cat cca gat ggc agt tgg aca cgt ttt acg ccg ttg cac His Ser Gln His Pro Asp Gly Ser Trp Thr Arg Phe Thr Pro Leu His 3960 3965 3970 3975	32883
gcc tta ccg ata gaa tat acc cat ccc cgc gcc caa ctt gcg gat tta Ala Leu Pro Ile Glu Tyr Thr His Pro Arg Ala Gln Leu Ala Asp Leu 3980 3985 3990	32931
atg ggg gcc ggg ctg tcc gat tta gtg ctg att ggt ccc aaa agc gtg Met Gly Ala Gly Leu Ser Asp Leu Val Leu Ile Gly Pro Lys Ser Val 3995 4000 4005	32979
cgt ttg tat gcc aat aac cgt gat ggt ttt acc gaa gga cgg gat gtg Arg Leu Tyr Ala Asn Asn Arg Asp Gly Phe Thr Glu Gly Arg Asp Val 4010 4015 4020	33027
gtg caa tcc ggt ggt atc acc ctg ccg tta ccg ggc gcc gat gcg cgt Val Gln Ser Gly Gly Ile Thr Leu Pro Leu Pro Gly Ala Asp Ala Arg 4025 4030 4035	33075
aag tta gtg gcc ttt agc gac gta ctc ggt tca ggc caa gca cat ttg Lys Leu Val Ala Phe Ser Asp Val Leu Gly Ser Gly Gln Ala His Leu 4040 4045 4050 4055	33123
gtt gaa gtt agt gcg acg aaa gtc acc tgc tgg cca aat ctg gga cat Val Glu Val Ser Ala Thr Lys Val Thr Cys Trp Pro Asn Leu Gly His	33171

4060		4065	4070
ggc cgt ttt ggt cag	cca atc aca tto	Pro Gly Phe Ser	caa tee gee 33219
Gly Arg Phe Gly Gln	Pro Ile Thr Lev		Gln Ser Ala
4075	4080		1085
gcc aat ttt aat cct Ala Asn Phe Asn Pro 4090	gat cga gtt cat Asp Arg Val His 4095	c ctg gcc gat ctg s Leu Ala Asp Leu 4100	gac ggt agt 33267 Asp Gly Ser
ggt cct gcc gat ctg Gly Pro Ala Asp Leu 4105	att tat gtt cat Ile Tyr Val His 4110	get gac cat etg Ala Asp His Leu 4115	gat att ttc 33315 Asp Ile Phe
agc aat gaa agt ggt	aac ggt ttt gca	a caa cca ttc aca	ctc cgt ttt 33363
Ser Asn Glu Ser Gly	Asn Gly Phe Ala	a Gln Pro Phe Thr	Leu Arg Phe
4120	4125	4130	4135
cct gac ggc ctg cgt	Phe Asp Asp Thi	tgc cag cta caa	gtg gct gat 33411
Pro Asp Gly Leu Arg		Cys Gln Leu Gln	Val Ala Asp
4140		4145	4150
gta cag gga tta ggg	gtt gtc agc ctc	ı Ile Leu Ser Val	ccg cat atg 33459
Val Gln Gly Leu Gly	Val Val Ser Let		Pro His Met
4155	4160		1165
gcg cca cac cat tgg Ala Pro His His Trp 4170	cgc tgc gat cto Arg Cys Asp Leo 4175	g acc aac gcg aaa 1 Thr Asn Ala Lys 4180	ccg tgg tta 33507 Pro Trp Leu
ctc agt gaa atg aac Leu Ser Glu Met Asn 4185	aac aac atg gga Asn Asn Met Gly 4190	a gcc cat cac acc y Ala His His Thr 4195	ctg cat tac 33555 Leu His Tyr
cgt agc tcc gtc cag	ttt tgg ctg gai	gaa aaa gcc gca	gcc tta gct 33603
Arg Ser Ser Val Gln	Phe Trp Leu Asp	Glu Lys Ala Ala	Ala Leu Ala
4200	4205	4210	4215
acc gga caa aca ccg	Val Cys Tyr Lei	g ccc ttc ccg gtc	cat acc ctg 33651
Thr Gly Gln Thr Pro		1 Pro Phe Pro Val	His Thr Leu
4220		4225	4230
tgg caa aca gaa acc	gag gat gaa ato	e Ser Gly Asn Lys	tta gtg acc 33699
Trp Gln Thr Glu Thr	Glu Asp Glu Ile		Leu Val Thr
4235	4240		1245
act tta cgt tac gct Thr Leu Arg Tyr Ala 4250	cac ggc gcc tgc His Gly Ala Tr 4255	g gat gga cgt gag o Asp Gly Arg Glu 4260	cgg gaa ttt 33747 Arg Glu Phe
cgc ggc ttt ggc tat Arg Gly Phe Gly Tyr 4265	gtt gag cag aca Val Glu Gln Thu 4270	a gac agc cat caa Asp Ser His Gln 4275	ctg gct caa 33795 Leu Ala Gln
ggc aat gcg ccg gaa	cgt aca tca cco	g gca ctt acc aaa	aac tgg tat 33843
Gly Asn Ala Pro Glu	Arg Thr Ser Pro	o Ala Leu Thr Lys	Asn Trp Tyr
4280	4285	4290	4295
gcc acc gga atc cct Ala Thr Gly Ile Pro 4300	gag gta gac aat Glu Val Asp Asr	acg cta tct gcc Thr Leu Ser Ala 4305	ggg tat tgg 33891 Gly Tyr Trp 4310
cgc ggt gat acg cag	gct ttc act ggt	Phe Thr Pro His	ttt act ctc 33939
Arg Gly Asp Thr Gln	Ala Phe Thr Gly		Phe Thr Leu
4315	4320		325

tgg aaa gag ggc aaa gat gtt cca ctg aca ccg gaa gat gac cac aat Trp Lys Glu Gly Lys Asp Val Pro Leu Thr Pro Glu Asp Asp His Asn 4330 4335 4340	33987
ctg tac tgg tta aac cgg gca cta aaa ggt caa cca ctg cgt agt gaa Leu Tyr Trp Leu Asn Arg Ala Leu Lys Gly Gln Pro Leu Arg Ser Glu 4345 4350 4355	34035
ctc tac ggg cta gat ggc agc gca cag cag aag atc ccc tat aca gtg Leu Tyr Gly Leu Asp Gly Ser Ala Gln Gln Lys Ile Pro Tyr Thr Val 4360 4365 4370 4375	34083
act gaa too ogo oca caa gtg ogo caa tta caa gat aac act acc ott Thr Glu Ser Arg Pro Gln Val Arg Gln Leu Gln Asp Asn Thr Thr Leu 4380 4385 4390	34131
tcc ccg gtg ctc tgg gcc tca gtg gtg gaa agt cgt agt tat cac tat Ser Pro Val Leu Trp Ala Ser Val Val Glu Ser Arg Ser Tyr His Tyr 4395 4400 4405	34179
gaa cgt atc atc agc gat ccc caa tgc aat cag gat atc act ctg tcc Glu Arg Ile Ile Ser Asp Pro Gln Cys Asn Gln Asp Ile Thr Leu Ser 4410 4415 4420	34227
agt gac cta ttc ggg caa ccg ctg aaa cag gtt tca gtg caa tat ccc Ser Asp Leu Phe Gly Gln Pro Leu Lys Gln Val Ser Val Gln Tyr Pro 4425 4430 4435	34275
cgc cgc aat aaa cca aca acc aat ccg tat ccc gat aca cta cca gat Arg Arg Asn Lys Pro Thr Thr Asn Pro Tyr Pro Asp Thr Leu Pro Asp 4440 4445 4450 4450	
act ctg ttt gcc agc agt tat gac gac caa caa caa cta ttg cgg tta Thr Leu Phe Ala Ser Ser Tyr Asp Asp Gln Gln Gln Leu Leu Arg Leu 4460 4465 4470	34371
acc tac cag caa tcc agt tgg cat cat cta att gct aat gaa ctc aga Thr Tyr Gln Gln Ser Ser Trp His His Leu Ile Ala Asn Glu Leu Arg 4475 4480 4485	3 441 9
gtg tta gga tta ccg gat ggt aca cgc agt gat gct ttc act tac gat Val Leu Gly Leu Pro Asp Gly Thr Arg Ser Asp Ala Phe Thr Tyr Asp 4490 4495 4500	34467
gct aaa cac gtg cct gtt gat ggt tta aat ctg gaa gct cta tgt gct Ala Lys His Val Pro Val Asp Gly Leu Asn Leu Glu Ala Leu Cys Ala 4505 4510 4515	3 451 5
gaa aat agc ctg att gcc gat gat aaa cct cgc gaa tac ctc aac cag Glu Asn Ser Leu Ile Ala Asp Asp Lys Pro Arg Glu Tyr Leu Asn Glr 4520 4525 4530 4535	ı
caa cga acg ttc tat acc gat ggg aaa acc gat gga aaa aat cca acc Gln Arg Thr Phe Tyr Thr Asp Gly Lys Thr Asp Gly Lys Asn Pro Th 4540 4545 4550	g 34611
cca ctg aaa aca ccg aca cga cag gct tta atc gcc ttt acc gaa acg Pro Leu Lys Thr Pro Thr Arg Gln Ala Leu Ile Ala Phe Thr Glu Th 4555 4560 4565	g 34659 r
gcg gta tta acg gaa tct ctg tta tcc gca ttt gat ggc ggt atc acg Ala Val Leu Thr Glu Ser Leu Leu Ser Ala Phe Asp Gly Gly Ile Th 4570 4575 4580	g 34707 r

cca gat gaa tta ccc ggc ctt ctg aca caa gca gga tac caa caa gaa Pro Asp Glu Leu Pro Gly Leu Leu Thr Gln Ala Gly Tyr Gln Gln Glu 4585 4590 4595	34755
cct tat ctg ttc cca ctc agt ggc gaa aac caa gtc tgg gta gca cgc Pro Tyr Leu Phe Pro Leu Ser Gly Glu Asn Gln Val Trp Val Ala Arg 4600 4605 4610 4615	34803
aaa ggc tat acc gat tac gga act gag gta caa ttt tgg cgt cct gtc Lys Gly Tyr Thr Asp Tyr Gly Thr Glu Val Gln Phe Trp Arg Pro Val 4620 4625 4630	34851
gca caa cgt aac acc cag tta acc ggg aaa acg act cta aaa tgg gat Ala Gln Arg Asn Thr Gln Leu Thr Gly Lys Thr Thr Leu Lys Trp Asp 4635 4640 4645	34899
acc cac tac tgt gtc atc act caa acc caa gac gcg gct ggt ttg act Thr His Tyr Cys Val Ile Thr Gln Thr Gln Asp Ala Ala Gly Leu Thr 4650 4660	34947
gtc tca gcc aat tat gac tgg cgt ttt ctc aca cct atg caa ctg act Val Ser Ala Asn Tyr Asp Trp Arg Phe Leu Thr Pro Met Gln Leu Thr 4665 4670 4675	34995
gat atc aac gat aat gtg cat atc ata acc ttg gat gcg cta gga cgc Asp Ile Asn Asp Asn Val His Ile Ile Thr Leu Asp Ala Leu Gly Arg 4680 4685 4690 4695	35043
cct gtc act caa cgt ttc tgg gga atc gaa aat ggt gtg gca aca ggt Pro Val Thr Gln Arg Phe Trp Gly Ile Glu Asn Gly Val Ala Thr Gly 4700 4705 4710	35091
tac tet tea eea gaa gea aaa eea tte aet eea eea gte gat gte aat Tyr Ser Ser Pro Glu Ala Lys Pro Phe Thr Pro Pro Val Asp Val Asn 4715 4720 4725	35139
gct gcc att gct ctg acc gga cca ctc cct gtc gcg cag tgt ctg gtc Ala Ala Ile Ala Leu Thr Gly Pro Leu Pro Val Ala Gln Cys Leu Val 4730 4735 4740	35187
tat gcg ccg gac agt tgg atg ccg cta ttc ggt cag gaa acc ttc aac Tyr Ala Pro Asp Ser Trp Met Pro Leu Phe Gly Gln Glu Thr Phe Asn 4745 4750 4755	35235
aca tta acg cag gaa gag caa aag aca ctg cgt gat tta cgg att atc Thr Leu Thr Gln Glu Gln Lys Thr Leu Arg Asp Leu Arg Ile Ile 4760 4765 4770 4775	35283
aca gaa gat tgg cgt att tgc gca ctg gct cgc cgc cgt tgg cta caa Thr Glu Asp Trp Arg Ile Cys Ala Leu Ala Arg Arg Trp Leu Gln 4780 4785 4790	35331
agt caa aaa gcc ggc aca cca ttg gtt aag ctg tta acc aac agc atc Ser Gln Lys Ala Gly Thr Pro Leu Val Lys Leu Leu Thr Asn Ser Ile 4795 4800 4805	35379
ggt tta cct ccc cac aac ctc atg ctg gct acg gac cgt tat gac cgt Gly Leu Pro Pro His Asn Leu Met Leu Ala Thr Asp Arg Tyr Asp Arg 4810 4815 4820	35427
gat tot gaa cag caa att ogt caa caa gto goa tto agt gat ggt ttt Asp Ser Glu Gln Gln Ile Arg Gln Gln Val Ala Phe Ser Asp Gly Phe 4825 4830 4835	35475
ggc cgt ttg ttg caa gcg gct gtg cgg cat gag gca ggc gaa gcc tgg	35523

Gly Arg Leu Leu Gln Ala Ala Val Arg His Glu Ala Gly Glu Ala Trp 4840 4845 4850 4855	
caa cgt aac caa gac ggt tct ctg gtg aca aaa atg gaa gat acc aaa Gln Arg Asn Gln Asp Gly Ser Leu Val Thr Lys Met Glu Asp Thr Lys 4860 4865 4870	35571
acg cgc tgg gcg att acg gga cgc act gaa tat gac aat aag ggg cag Thr Arg Trp Ala Ile Thr Gly Arg Thr Glu Tyr Asp Asn Lys Gly Gln 4875 4880 4885	35619
gcg ata cga act tat cag ccc tat ttc ctc aat gac tgg cga tat gtg Ala Ile Arg Thr Tyr Gln Pro Tyr Phe Leu Asn Asp Trp Arg Tyr Val 4890 4895 4900	35667
agt gat gac agc gcc aga aaa gag gcc tat gcc gat act cat atc tat Ser Asp Asp Ser Ala Arg Lys Glu Ala Tyr Ala Asp Thr His Ile Tyr 4905 4910 4915	35715
gat ccg att ggg cgg gaa atc caa gtt atc acg gca aaa ggc tgg ctg Asp Pro Ile Gly Arg Glu Ile Gln Val Ile Thr Ala Lys Gly Trp Leu 4920 4925 4930 4935	35763
cgg cag aac caa tat ttc ccg tgg ttt acc gtg agt gaa gat gaa aat Arg Gln Asn Gln Tyr Phe Pro Trp Phe Thr Val Ser Glu Asp Glu Asn 4940 4945 4950	35811
gat ttg tcc gct gac gcg ctc gtg taa ttgaatcaag attcgctcgt Asp Leu Ser Ala Asp Ala Leu Val 4955 4960	35858
ttaatgttaa cgagcgaata taatatacct aatagatttc gagttgcagc gcggcggc	aa 35918
gtgaacgaat ccccaggagc atagataact atgtgactgg ggtgagtgaa agcagcca	ac 35978
aaagcagcag cttgaaagat gaagggtata aataagaaac tgcattgtga gttctaaa	ta 36038
gagtagcagc atattttatt gccttttatt tcataggtaa taaaattcaa ttgctgta	aa 36098
aatetgteat eatgagaaet aaaaataaea aetttetett etgeaagaga aateaata	at 36158
tcaattaaaa atgttataga atctgaatca agaccatttg ttggctcatc aaaaatat	aa 36218
acatccgcat cggtaataaa agctgatgtc aatagaaatt tcttttttat cccaagtg	ac 36278
atatgtccat actcaatacc agaataatta gatataccaa aaccatttaa atagtaat	ct 36338'
aattgatatt ttaaattact tttcctataa cgctgactta aattaatcac atccattc	
gtgatgaaat tataaaagtt aacattatcc gatagataaa aaccatgctg ttgcaaat	ta 36458
aateggetet titeteeett tittataaaa tiaaeeatte etitiitaae etiatita	aca 36518
ccagcaatac ttgaaagaaa agtcgtttta cccgccccat taactcccgc aatacggt	tt 36578
aatccaaccc gaaaatcaca attgactcct gaaaaaatag tettaccatt aataacaa	acc 36638
totaacccaa taacttcaag cataaataac cootaaaaat aacgtaaaaa agaaaata	aac 36698
accaacaata ataattttcg tgtattgcgt tctcaacaga gaaatagaag aaacaata	aat 36758
agaaagaaaa gcataagata aaaatataat cacaggaaaa gatttaacaa caagaaa	gca 36818
aaaaataaaa aaacaaagca aataaaaaaa caaagaaata ccataattaa aaaagaa	tat 36878

WO 99/42589

tttccgcaca gataaaaagt tggacaaata tgaaagataa tttatttcaa tatatgatag 36938 attataaaat aacaacatgc atatatataa aacaacactg gcatatatta atgatatata 36998 atcaqcettq tttgggattt gagaaaagge actttcacat aatagatata aaagcaqaac 37058 agataatgcc ataataagca cagacatttt atttttatta aaacaataac gaagattcat 37118 tatataaggc aatgaaaaaa aacctgatga aaataatttt ttatttctat taattatata 37178 acatggtgtg aaattcaaat ataatatcaa tgctactaat ggaataacta atgtaaaaat 37238 caaatcatat aatattccac tcctgaatga tgccgccaga agaaagaaca cagcaacaat 37298 aaaaaaatgc aaaaaactta attcaaataa gcaaaatcca attacagcaa aagaaactat 37358 caaaaaaaac acagatgaaa ggtaatgcaa ataattaaca ttttcgtaaa aaaaacctat 37418 aaagaagaaa ataactateg gaaaagcact ataaataaaa aaaacgatac gactaaaaaa 37478 caacgttttt ttacctacca aagaaacgat gattgaattc tcctttgcag aaggaaaaaa 37538 tgattttatg tagccatctg gaataataat attggaagat aaagttatta aaacctcaaa 37658 gataccactg aactttgccg gaagtaataa aagaaaaagg aatataatga catttttatt 37718 cccagacgca aatttetta tectacettt atattecaag geateagega ttattaaatt 37778 catactgcct ctctaaaacc aaaatctaaa taatgtcctt ggtgaatctt tagggaattt 37838 cgtcctggaa tgcaaatata aatagttact gaaaacaata cattgatttt taattaaata 37898 ctggcgatat gaccttaatg atgctacttt attttccagt attcaattcg 37948

<210> 12

<211> 954

<212> PRT

<213> Photorhabdus luminescens

<400> 12

Met Lys Asn Ile Asp Pro Lys Leu Tyr Gln Lys Thr Pro Val Val Asn 1 5 10

Ile Tyr Asp Asn Arg Gly Leu Thr Ile Arg Asn Ile Asp Phe His Arg 20 25 30

Thr Thr Ala Asn Gly Asp Thr Asp Ile Arg Ile Thr Arg His Gln Tyr 35 40 45

Asp Ser Leu Gly His Leu Ser Gln Ser Thr Asp Pro Arg Leu Tyr Glu 50 60

Ala Lys Gln Lys Ser Asn Phe Leu Trp Gln Tyr Asp Leu Thr Gly Asn 65 70 75

Ile Leu Cys Thr Glu Ser Val Asp Ala Gly Arg Thr Val Thr Leu Asn 85 90 95

Asp Ile Glu Gly Arg Pro Leu Leu Thr Val Thr Ala Thr Gly Val Ile 100 105 110

Cln Thr Arg Gln Tyr Glu Thr Ser Ser Leu Pro Gly Arg Leu Leu Ser

		115					120					125			
Val	Thr 130	Glu	Gln	Ile	Pro	Glu 135	Lys	Thr	Ser	Arg	Ile 140	Thr	Glu	Arg	Leu
Ile 145	Trp	Ala	Gly	Asn	Ser 150	Glu	Ala	Glu	Lys	Asn 155	His	Asn	Leu	Ala	Ser 160
Gln	Cys	Val	Arg	His 165	Tyr	Asp	Thr	Ala	Gly 170	Val	Thr	Arg	Leu	Glu 175	Ser
Leu	Ser	Leu	Thr 180		Thr	Val	Leu	Ser 185	Gln	Ser	Ser	Gln	Leu 190	Leu	Ser
Asp	Thr	Glr 195	Glu	Ala	Ser	Trp	Thr 200	Gly	Asp	Asn	Glu	Thr 205	Val	Trp	Gln
Asn	Met 210		ı Ala	Asp	Asp	Ile 215	Tyr	Thr	Thr	Leu	Ser 220	Ala	Phe	Asp	Ala
Thr 225		' Alá	i Leu	Leu	Thr 230		Thr	Asp	Ala	Lys 235		Asn	Ile	Gln	Arg 240
Leu	Thr	Туз	: Asp	Val 245		Gly	Gln	Leu	Asn 250		Ser	Trp	Leu	Thr 255	Leu
Lys	Asp	Glr	n Pro 260		Gln	Val	. Ile	1le 265		Ser	Leu	Thr	Туг 270	Ser	Ala
Ala	Gly	ر Gl 27!		s Leu	ı Arg	Glu	280		Gly	Asr.	Gly	Val 285		Thr	Glu
Tyr	Se: 29	-	r Glu	ı Pro	Glu	Th: 295		ı Glr	ı Lev	ı Ile	300 300		Lys	Thr	His
Arg 305	-	o Se	r Ası	o Ala	a Lys 310		l Lev	ı Glr	n Asp	315		Ty1	Glu	ı Tyr	320
Pro	o Va	l Gl	y Asi	n Vai		e Sei	r Ile	e Arg	g Ası 330		o Ala	a Glu	ı Ala	Thi 335	Arg
Phe	e Tr	p Hi	s As: 34		n Ly:	s Va	l Ala	a Pro 34:		u Ası	n Thi	c Ty:	7hi 350	с Тул Э	Asp
Se	r Le	u Ty 35		n Le	u Il	e Se	r Ala 36		r Gly	y Ar	g Glu	и Ме 36	t Ala 5	a Ast	ı Ile
Gl	y Gl 37		n Se	r As	n Gl	n Le 37	u Pro 5	o Se	r Le	u Th	r Le 38	u Pr O	o Se:	r Ası	o Asn
As 38	_	r T	nr Th	ır As	n Ty 39		r Ar	g Th	r Ty	r Th		r As	p Ar	g Gl	y Gly 400
As	n Le	eu Th	ır Ly	s Il 40		n Hi	s Se	r Se	r Pr 41		a Th	r Gl	n As	n As 41	n Tyr 5
าใก	r Ti	nr As	sn Il 42		ır Va	l Se	er As	n Ar 42		r As	n Ar	g Al	a Va 43	1 Le 0	u Ser
Th	r L		hr G] 35	lu As	sp Pr	CA O	la Gl 44		al As	ap Al	a Le	eu Ph 44		p Al	a Gly
Gl	-	is G 50	ln As	sn Th	r Le		le Se 55	er Gl	ly Gl	ln As	n Le 46		n Tr	p As	n Thi

Arg 465	Gly	Glu	Leu	Gln	His 470	Val	Thr	Leu	Val	Lys 475	Arg	Asp	Lys	Gly	Ala 480
Asn	Asp	Asp	Arg	Glu 485	Trp	Tyr	Arg	Tyr	Ser 490	Ser	Asp	Gly	Arg	Arg 495	Ile
Leu	Lys	Ile	Asn 500	Glu	Gln	Gln	Thr	Ser 505	Ser	Asn	Ser	Gln	Thr 510	Gln	Arg
Ile	Thr	Тут 515	Leu	Pro	Ser	Leu	Glu 520	Leu	Arg	Leu	Thr	Gln 525	Asn	Ser	Thr
Ile	Thr 530	Thr	Glu	Asp	Leu	Gln 535	Val	Ile	Thr	Val	Gly 540	Glu	Ala	Gly	Arg
Ala 545	Gln	Val	Arg	Val	Leu 550	His	Trp	Asp	Ser	Gly 5 5 5	Gln	Pro	Glu	Asp	11e 560
Asp	Asn	Asn	Gln	Leu 565	Arg	Tyr	Ser	Tyr	Asp 570	Asn	Leu	Ile	Gly	Ser 575	Ser
Gln	Leu	Glu	Leu 580	Asp	Ser	Lys	Gly	Glu 585	Ile	Ile	Ser	Glu	Glu 590	Glu	Tyr
Tyr	Pro	Tyr 595	Gly	Gly	Thr	Ala	Leu 600	Trp	Ala	Thr	Arg	Lys 605	Arg	Thr	Glu
Ala	Ser 610	Tyr	Lys	Thr	Ile	Arg 615	Tyr	Ser	Gly	Lys	Glu 620	Arg	Asp	Ala	Thr
Gly 625	Leu	Tyr	Tyr	Tyr	Gly 630	Tyr	Arg	Tyr	Tyr	Gln 635	Pro	Trp	Val	Gly	Arg 640
Trp	Leu	Ser	Ala	Asp 645	Pro	Ala	Gly	Thr	Val 650		Gly	Leu	Asn	Leu 655	Tyr
Arg	Met	Val	Arg 660	Asn	Asn	Pro	Val	Thr 665	Leu	Leu	Asp	Pro	Asp 670	Gly	Leu
Met	Pro	Thr 675	Ile	Ala	Glu	Arg	11e 680		Ala	Leu	Gln	Lys 685	Asn	Lys	Val
Ala	Asp 690		Ala	Pro	Ser	Pro 695		Asn	Ala	Thr	Asn 700		Ala	Ile	Asn
Ile 705		Pro	Pro	Val	Ala 710		Lys	Pro	Thr	Leu 715		Lys	Ala	Ser	Thr 720
Ser	Ser	Gln	Ser	Thr 725		Tyr	Pro	Ile	Lys 730		Ala	Ser	Ile	Lys 735	Pro
Thr	Thr	Ser	Gly 740		Ser	IJ€	Thr	· Ala 745		Leu	Ser	Pro	Val 750		Asn
Lys	Ser	Thr 755	Pro	Glu	Il∈	: Ser	Leu 760		Glu	Ser	Thr	Gln 765		Asn	Ser
Ser	Ser 770		lle	Ser	Thr	Asn 775		Gln	Lys	Lys	Ser 780		Thr	Leu	Tyr
Arg 785		Asp) Asn	Arg	Ser 790		e Glu	ı Asp	Met	Gln 795		: Lys	Phe	Pro	Glu 800
Gly	, by	Lys	s Ala	Trp 805		Pro	Leu	ı Asp	Thr 810		: Met	Ala	Arg	Gln 815	

WO 99/42589 PCT/EP99/01015

Ala Ser Val Phe Ile Gly Gln Lys Asp Thr Ser Asn Leu Pro Lys Glu 820 825 830

Thr Val Lys Asn Ile Asn Thr Trp Gly Thr Lys Pro Lys Leu Asn Asp 835 840 845

Leu Ser Thr Tyr Ile Lys Tyr Thr Lys Asp Lys Ser Thr Val Trp Val 850 855 860

Ser Thr Ala Ile Asn Thr Glu Ala Gly Gly Gln Ser Ser Gly Ala Pro 865 870 875 880

Leu His Glu Ile Asn Met Asp Leu Tyr Glu Phe Thr Ile Asp Gly Gln 885 890 895

Lys Leu Asn Pro Leu Pro Arg Gly Arg Ser Lys Asp Arg Val Pro Ser 900 905 910

Leu Leu Leu Asp Thr Pro Glu Ile Glu Thr Ala Ser Ile Ile Ala Leu 915 920 925

Asn His Gly Pro Val Asn Asp Ala Glu Val Ser Phe Leu Thr Thr Ile 930 935 940

Pro Leu Lys Asn Val Lys Pro Tyr Lys Arg 945 950

<210> 13

<211> 2522

<212> PRT

<213> Photorhabdus luminescens

<400> 13

Met Ile Leu Lys Gly Ile Asn Met Asn Ser Pro Val Lys Glu Ile Pro 1 5 10 15

Asp Val Leu Lys Ile Gln Cys Gly Phe Gln Cys Leu Thr Asp Ile Ser 20 25 30

His Ser Ser Phe Asn Glu Phe His Gln Gln Val Ser Glu His Leu Ser 35 40 45

Trp Ser Glu Ala His Asp Leu Tyr His Asp Ala Gln Gln Ala Gln Lys 50 60

Asp Asn Arg Leu Tyr Glu Ala Arg Ile Leu Lys Arg Thr Asn Pro Gln 65 70 75 80

Leu Gln Asn Ala Val His Leu Ala Ile Val Ala Pro Asn Ala Glu Leu 85 90 95

Ile Gly Tyr Asn Asn Gln Phe Ser Gly Arg Ala Ser Gln Tyr Val Ala
100 105 110

Pro Gly Thr Val Ser Ser Met Phe Ser Pro Ala Ala Tyr Leu Thr Glu 115 120 125

Leu Tyr Arg Glu Ala Arg Asn Leu His Ala Ser Asp Ser Val Tyr Arg 130 135 140

Leu Asp Thr Arg Arg Pro Asp Leu Lys Ser Met Ala Leu Ser Gln Gln 145 150 155 160

Asn	Met	Asp	Thr	Glu 165	Leu	Ser	Thr	Leu	Ser 170	Leu	Ser	Asn	Glu	Leu 175	Leu
Leu	Glu	Ser	Ile 180	Lys	Thr	Glu	Ser	Lys 185	Leu	Asp	Asn	Tyr	Thr 190	Gln	Val
Met	Glu	Met 195	Leu	Ser	Ala	Phe	Arg 200	Pro	Ser	Gly	Ala	Thr 205	Pro	Tyr	His
Asp	Ala 210	Tyr	Glu	Asn	Val	Arg 215	Lys	Val	Ile	Gln	Leu 220	Gln	Asp	Pro	Gly
Leu 225	Glu	Gln	Leu	Asn	Ala 230	Ser	Pro	Ala	Ile	Ala 235	Gly	Leu	Met	His	Gln 240
Ala	Ser	Leu	Leu	Gly 2 4 5	Ile	Asn	Ala	Ser	Ile 250	Ser	Pro	Glu	Leu	Phe 255	Asn
Ile	Leu	Thr	Glu 260	Glu	Ile	Thr	Glu	Gly 265	Asn	Ala	Glu	Glu	Leu 270	Tyr	Lys
Lys	Asn	Phe 275	Gly	Asn	Ile	Glu	Pro 280	Ala	Ser	Leu	Ala	Met 285	Pro	Glu	Tyr
Leu	Arg 290	Arg	Tyr	Tyr	Asn	Leu 295	Ser	Asp	Glu	Glu	Leu 300	Ser	Gln	Phe	Ile
Gly 305	Lys	Ala	Ser	Asn	Phe 310	Gly	Gln	Gln	Glu	Туг 315		Asn	Asn	Gln	Leu 320
Ile	Thr	Pro	Ile	Val 325	Asn	Ser	Asn	Asp	Gly 330		Val	Lys	Val	Тут 335	Arg
Ile	Thr	Arg	Glu 340	_	Thr	Thr	Asn	Ala 345		Gln	Val	Asp	Val 350	Glu	Leu
Phe	Pro	Tyr 355	Gly	Gly	Glນ	Asn	Туг 360		Leu	Asn	Tyr	Lys 365		Lys	Asp
Ser	Arg 370		Asp	Val	Ser	Тут 375		Ser	Ile	Lys	Leu 380		Asp	Lys	Arg
Glu 385		Ile	Arg	Il∈	Glu 390	_	Ala	Pro	Gln	Val 395		Ile	Glu	Tyr	Ser 400
Glu	His	Ile	Thr	Leu 405		Thr	Thr	Asp	1l∈ 410		Gln	Pro	Ph∈	Glu 415	
Gly	Leu	Thr	Arg 420		Тут	Pro	Ser	Ser 425		Tr	Ala	Tyr	Ala 430		Ala
Lys	: Phe	Thr 435	Ile	Glu	Glu	Tyr	440		Тут	Ser	Phe	Leu 445		Lys	Leu
Asn	1 Lys 450		ıl∈	e Arc	Leu	Ser 455		, Ala	Thr	Glu	1 Leu 460		Pro	Thr	Ile
Leu 465		ı Ser	Ile	e Val	Arc 470		· Val	. Asr	Glr	1 Glr 475	_	Asp	ıle	e Asn	Ala 480
Glu	ı Val	Leu	ı Gly	485		l Phe	e Leu	ı Thr	Lys 490		туг	. Met	Glr	Arc 495	
Alā	a Ile	e Ast	n Ala 500		ı Thi	Ala	a Leu	1 Ile 509		ı Cys	s Asr	n Ala	Le. 51(e Ser

Gln Arg Ser Tyr Asp Asn Gln Pro Ser Gln Phe Asp Arg Leu Phe Asn 520 Thr Pro Leu Leu Asn Gly Gln Tyr Phe Ser Thr Gly Asp Glu Glu Ile Asp Leu Asn Pro Gly Ser Thr Gly Asp Trp Arg Lys Ser Val Leu Lys 555 Arg Ala Phe Asn Ile Asp Asp Ile Ser Leu Tyr Arg Leu Lys Ile 565 570 575 Thr Asn His Asn Asn Gln Asp Gly Lys Ile Lys Asn Asn Leu Asn Asn Leu Ser Asp Leu Tyr Ile Gly Lys Leu Leu Ala Glu Ile His Gln Leu Thr Ile Asp Glu Leu Asp Leu Leu Leu Val Ala Val Gly Glu Gly Glu Thr Asn Leu Ser Ala Ile Ser Asp Lys Gln Leu Ala Ala Leu Ile Arg Lys Leu Asn Thr Ile Thr Val Trp Leu Gln Thr Gln Lys Trp Ser Ala 650 Phe Gln Leu Phe Val Met Thr Ser Thr Ser Tyr Asn Lys Thr Leu Thr 665 Pro Glu Ile Lys Asn Leu Leu Asp Thr Val Tyr His Gly Leu Gln Gly Phe Asp Lys Asp Lys Ala Asn Leu Leu His Val Met Ala Pro Tyr Ile Ala Ala Thr Leu Gln Leu Ser Ser Glu Asn Val Ala His Ser Val Leu Leu Trp Ala Asp Lys Leu Lys Pro Gly Asp Gly Ala Met Thr Ala Glu 725 730 735 Lys Phe Trp Asp Trp Leu Asn Thr Gln Tyr Thr Pro Asp Ser Ser Glu 740 745 750 Val Leu Ala Thr Gln Glu His Ile Val Gln Tyr Cys Gln Ala Leu Ala Gln Leu Glu Met Val Tyr His Ser Thr Gly Ile Asn Glu Asn Ala Phe Arg Leu Phe Val Thr Lys Pro Glu Met Phe Gly Ser Ser Thr Glu Ala 795 Val Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr Arg Phe Ala Asp Trp Val Asn Ala Leu Gly Glu Lys Ala Ser Ser Val Leu Ala Ala Phe Glu Ala Asn Ser Leu Thr Ala Glu Gln Leu Ala Asp Ala Met Asn

Leu Asp Ala Asn Leu Leu Gln Ala Ser Thr Gln Ala Gln Asn His

	850					855					860				
Gln 865	His	Leu	Pro	Pro	Val 870	Thr	Gln	Lys	Asn	Ala 875	Phe	Ser	Cys	Trp	Thr 880
Ser	Ile	Asp	Thr	Ile 885	Leu	Gln	Trp	Val	Asn 890	Val	Ala	Gln	Gln	Leu 895	Asn
Val	Ala	Pro	Gln 900	Gly	Val	Ser	Ala	Leu 905	Val	Gly	Leu	Asp	Тут 910	Ile	Gln
Leu	Asn	Gln 915	Lys	Ile	Pro	Thr	Тут 920	Ala	Gln	Trp	Glu	Ser 925	Ala	Gly	Glu
Ile	Leu 930	Thr	Ala	Gly	Leu	Asn 935	Ser	Gln	Gln	Ala	Asp 940	Ile	Leu	His	Ala
Phe 945	Leu	Asp	Glu	Ser	Arg 950	Ser	Ala	Ala	Leu	Ser 955	Thr	Tyr	Tyr	Ile	Arg 960
Gln	Val	Ala	Lys	Pro 965	Ala	Ala	Ala	Ile	Lys 970	Ser	Arg	Asp	Asp	Leu 975	Tyr
Gln	Tyr	Leu	Leu 980	Ile	Asp	Asn	Gln	Val 985	Ser	Ala	Ala	Ile	Lys 990	Thr	Thr
Arg	Ile	Ala 995	Glu	Ala	Ile		Ser 1000	Ile	Gln	Leu	_	Val 1005	Asn	Arg	Thr
	Glu 1010		Val	Glu		Asn 1015	Ala	His	Ser	_	Val 1020		Ser	Arg	Gln
Phe 025		Ile	Asp	_	Asp 1030		Tyr	Asn		Arg 1035	_	Ser	Thr		Ala 1040
Gly	Val	Ser		Leu 1045		Tyr	Tyr	Pro	Glu 1050		Tyr	Ile		Pro 1055	Thr
Met	Arg		Gly 1060		Thr	Lys	Met	Met 1065		Ala	Leu		Gln 1070	Ser	Val
Ser	Gln	Ser 1075		Leu	Asn	Ala	Asp 1080		Val	Glu	Asp	Ala 1085		Met	Ser
Tyr	Leu 1090		Ser	Phe	Glu	Glr. 1095		Ala	Asr	Leu	Lys 1100		Ile	Ser	Ala
Тут 105		: Asp) Asn	lle	1110	Asr	Asp	Glr	Gly	' Leu 1115	Thr	Tyr	Phe	Ile	Gly 1120
Leu	Ser	Glu	Thr	1125		Gly	/ Glu	туг	тут 1130		Arg	g Ser	· Val	Asp 1135	His
Ser	Lys	Ph∈	Ser 1140		Gly	Lys	s Phe	Ala 114		a Asr	n Alá	a Trp	Ser 1150		qrr
His	: Lys	1155		Cys	Pro) Ile	2 Asr 1160		тут	r Arg	g Sei	Thr 1169		e Arg	Pro
Va]	Met 117(-	. Lys	s Sei	: Arg	1179		Lei	ı Lei	TTT	1180	_	ı Glr	Lys	Glu
11e		Ly:	s Glr	n Thu	Gly 1190		n Sei	. Ly:	s Asq	o Gly 1199		c Glr	Thr	Glu	Thr 1200

- Asp Tyr Arg Tyr Glu Leu Lys Leu Ala His Ile Arg Tyr Asp Gly Thr 1205 1210 1215
- Trp Asn Thr Pro Ile Thr Phe Asp Val Asn Glu Lys Ile Ser Lys Leu 1220 1225 1230
- Glu Leu Ala Lys Asn Lys Ala Pro Gly Leu Tyr Cys Ala Gly Tyr Gln 1235 1240 1245
- Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Asn Gln Gln Asp Thr Leu 1250 1255 1260
- Asp Ser Tyr Lys Thr Ala Ser Met Gln Gly Leu Tyr Ile Phe Ala Asp 265 1270 1275 1280
- Met Glu Tyr Lys Asp Met Thr Asp Gly Gln Tyr Lys Ser Tyr Arg Asp 1285 1290 1295
- Asn Ser Tyr Lys Gln Phe Asp Thr Asn Ser Val Arg Arg Val Asn Asn 1300 1305 1310
- Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Asn Ser Arg Lys 1315 1320 1325
- Gly Tyr Asp Trp Gly Asp Tyr Tyr Leu Ser Met Val Tyr Asn Gly Asp 1330 1335 1340
- Ile Pro Thr Ile Ser Tyr Lys Ala Thr Ser Ser Asp Leu Lys Ile Tyr 345 1350 1355 1360
- Ile Ser Pro Lys Leu Arg Ile Ile His Asn Gly Tyr Glu Gly Gln Gln 1365 1370 1375
- Arg Asn Gln Cys Asn Leu Met Asn Lys Tyr Gly Lys Leu Gly Asp Lys 1380 1385 1390
- Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn Ser Ser Asn 1395 1400 1405
- Lys Leu Met Phe Tyr Pro Val Tyr Gln Tyr Asn Gly Asn Val Ser Gly 1410 1415 1420
- Leu Ser Gln Gly Arg Leu Leu Phe His Arg Asp Thr Asn Tyr Ser Ser 425 1430 1435 1440
- Lys Val Glu Ala Trp Ile Pro Gly Ala Gly Arg Ser Leu Thr Asn Pro 1445 1450 1455
- Asn Ala Ala Ile Gly Asp Asp Tyr Ala Thr Asp Ser Leu Asn Lys Pro 1460 1465 1470
- Asn Asp Leu Lys Gln Tyr Val Tyr Met Thr Asp Ser Lys Gly Thr Ala 1475 1480 1485
- Thr Asp Val Ser Gly Pro Val Asp Ile Asn Thr Ala Ile Ser Pro Ala 1490 1495 1500
- Lys Val Gln Val Thr Val Lys Ala Gly Ser Lys Glu Gln Thr Phe Thr 505 1510 1515 1520
- Ala Asp Lys Asn Val Ser Ile Gln Pro Ser Pro Ser Phe Asp Glu Met 1525 1530 1535
- Asn Tyr Gln Phe Asn Ala Leu Glu Ile Asp Gly Ser Ser Leu Asn Phe 1540 1545 1550

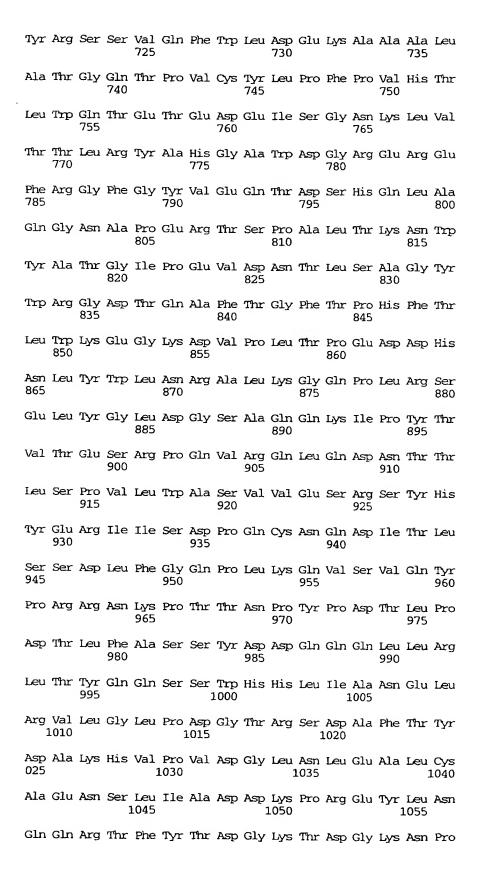
- Thr Asn Asn Ser Ala Ser Ile Asp Ile Thr Phe Thr Ala Phe Ala Glu 1555 1560 1565
- Asp Gly Arg Lys Leu Gly Tyr Glu Ser Phe Ser Ile Pro Ile Thr Arg 1570 1575 1580
- Lys Val Ser Thr Asp Asn Ser Leu Thr Leu Arg His Asn Glu Asn Gly 585 1590 1595 1600
- Ala Gln Tyr Met Gln Trp Gly Val Tyr Arg Ile Arg Leu Asn Thr Leu 1605 1610 1615
- Phe Ala Arg Gln Leu Val Ala Arg Ala Thr Thr Gly Ile Asp Thr Ile 1620 1625 1630
- Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly Lys Gly 1635 1640 1645
- Phe Tyr Ala Thr Phe Val Ile Pro Pro Tyr Asn Pro Ser Thr His Gly 1650 1655 1660
- Asp Glu Arg Trp Phe Lys Leu Tyr Ile Lys His Val Val Asp Asn Asn 665 1670 1675 1680
- Ser His Ile Ile Tyr Ser Gly Gln Leu Lys Asp Thr Asn Ile Ser Thr 1685 1690 1695
- Thr Leu Phe Ile Pro Leu Asp Asp Val Pro Leu Asn Gln Asp Tyr Ser 1700 1705 1710
- Ala Lys Val Tyr Met Thr Phe Lys Lys Ser Pro Ser Asp Gly Thr Trp 1715 1720 1725
- Trp Gly Pro His Phe Val Arg Asp Asp Lys Gly Ile Val Thr Ile Asn 1730 1735 1740
- Pro Lys Ser Ile Leu Thr His Phe Glu Ser Val Asn Val Leu Asn Asn 745 1750 1755 1760
- Ile Ser Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ser Leu Tyr Phe 1765 1770 1775
- Trp Glu Leu Phe Tyr Tyr Thr Pro Met Leu Val Ala Gln Arg Leu Leu 1780 1785 1790
- His Glu Gln Asn Phe Asp Glu Ala Asn Arg Trp Leu Lys Tyr Val Trp 1795 1800 1805
- Ser Pro Ser Gly Tyr Ile Val His Gly Gln Ile Gln Asn Tyr Gln Trp 1810 1815 1820
- Asn Val Arg Pro Leu Leu Glu Asp Thr Ser Trp Asn Ser Asp Pro Leu 825 1830 1835 1840
- Asp Ser Val Asp Pro Asp Ala Val Ala Gln His Asp Pro Met His Tyr 1845 1850 1855
- Lys Val Ser Thr Phe Met Arg Thr Leu Asp Leu Leu Ile Ala Arg Gly 1860 1865 1870
- Asp His Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Asn Glu Ala Lys 1875 1880 1885
- Met Trp Tyr Met Gln Ala Leu His Leu Leu Gly Asp Lys Pro Tyr Leu

1890		1	. 89 5		1900		
Pro Leu : 905	Ser Thr	Thr Trp 1910	Asn Asp	Pro Arg	Leu Asp 1915	Lys Ala	Ala Asp 1920
Ile Thr		Ser Ala 1925	His Ser	Ser Ser 1930	Ile Val	Ala Leu 1	Arg Gln .935
Ser Thr	Pro Ala 1940	Leu Leu		Arg Ser 1945	Ala Asn	Thr Leu 1950	Thr Asp
	Leu Pro .955	Gln Ile	Asn Glu 1960		Met Asn	Tyr Trp 1965	Gln Thr
Leu Ala 1970	Gln Arg		Asn Leu 1975	Arg His	Asn Leu 1980	Ser Ile	Asp Gly
Gln Pro 985	Leu Tyr	Leu Pro 1990		Ala Thr	Pro Ala 1995	Asp Pro	Lys Ala 2000
Leu Leu		Ala Val 2005	Ala Thr	Ser Glr 2010		Gly Lys	Leu Pro 2015
	2020			2025		Leu Glu 2030	
	Met Val 2035	. Ser Glr	Leu Thr 2040		e Gly Ser	Thr Leu 2045	Gln Asn
Ile Ile 2050		ı Gln Asp	Ala Glu 2055	ı Ala Lei	a Asn Ala 2060	a Leu Leu)	Gln Asn
Gln Ala 065	Ala Glu	Leu Ile 2070		r Asn Le	2075 2075	e Gln Asp	Lys Thr 2080
Ile Glu	Glu Le	1 Asp Ala 2085	a Glu Ly	s Thr Va 209	l Leu Gli 0	ı Lys Ser	Lys Ala 2095
Gly Ala	Gln Ser 210		e Asp Se	r Tyr Se 2105	r Lys Le	u His Asq 211(o Glu Asn)
	2115		212	0		2125	r Ala Ala
2130)		2135		214	0	a Ala Ala
Asp Lei 145	ı Val Pr	o Asn Il 215		y Phe Al	a Gly Gl 2155	y Gly Se:	r Arg Trp 2160
Gly Ala	a Ile Al	a Glu Al 2165	a Thr Gl	y Tyr Va 217		u Phe Se	r Ala Asn 2175
Val Me	t Asn Tr 218		a Asp Ly	s Ile Se 2185	er Gln Se	er Glu Th 219	r Tyr Arg O
Arg Ar	g Arg G] 2195	n Glu Tr	p Glu II 220		rg Asn As	sn Ala Gl 2205	u Ala Glu
Leu Ly 221		eu Asp Al	a Gln La 2215	eu Lys S	er Leu Al 222		g Arg Glu
Ala Al 225	a Val Le	eu Gln Ly 22:		er Leu L	ys Thr Gi 2235	ln Gln Gl	u Gln Thr 2240

- Gln Ala Gln Leu Ala Phe Leu Gln Arg Lys Phe Ser Asn Gln Ala Leu 2245 2250 2255
- Tyr Asn Trp Leu Arg Gly Arg Leu Ala Ala Ile Tyr Phe Gln Phe Tyr 2260 2265 2270
- Asp Leu Ala Ile Ala Arg Cys Leu Met Ala Glu Gln Ala Tyr Arg Trp 2275 2280 2285
- Glu Ile Ser Asp Asp Ser Ala Arg Phe Ile Lys Pro Gly Ala Trp Gln 2290 2295 2300
- Gly Thr Tyr Ala Gly Leu Leu Ala Gly Glu Thr Leu Met Leu Ser Leu 305 2310 2315 2320
- Ala Gln Met Glu Asp Ala His Leu Arg Arg Asp Lys Arg Ala Leu Glu 2325 2330 2335
- Val Glu Arg Thr Val Ser Leu Ala Glu Ile Tyr Ala Gly Leu Pro Gln 2340 2345 2350
- Asp Lys Gly Pro Phe Ser Leu Thr Gln Glu Ile Glu Lys Leu Val Asn 2355 2360 2365
- Ala Gly Ser Gly Ser Ala Gly Ser Gly Asn Asn Leu Ala Phe Gly 2370 2375 2380
- Ala Gly Thr Asp Thr Lys Thr Ser Leu Gln Ala Ser Ile Ser Leu Ala 385 2390 2395 2400
- Asp Leu Lys Ile Arg Glu Asp Tyr Pro Glu Ser Ile Gly Lys Ile Arg 2405 2410 2415
- Arg Ile Lys Gln Ile Ser Val Thr Leu Pro Ala Leu Leu Gly Pro Tyr 2420 2425 2430
- Gln Asp Val Gln Ala Ile Leu Ser Tyr Gly Asp Lys Ala Gly Leu Ala 2435 2440 2445
- Asn Gly Cys Ala Ala Leu Ala Val Ser His Gly Thr Asn Asp Ser Gly 2450 2455 2460
- Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys Phe Leu Pro Phe Glu Gly 465 2470 2475 2480
- Ile Ala Ile Asp Gln Gly Thr Leu Thr Leu Ser Phe Pro Asn Ala Ser 2485 2490 2495
- Thr Pro Ala Lys Gly Lys Gln Ala Thr Met Leu Lys Thr Leu Asn Asp 2500 2505 2510
- Ile Ile Leu His Ile Arg Tyr Thr Ile Lys 2515 2520
- <210> 14
- <211> 1481
- <212> PRT
- <213> Photorhabdus luminescens
- <400> 14
- Met Gln Asn Ser Gln Thr Phe Ser Met Thr Glu Leu Ser Leu Pro Lys 1 5 10 15
- Gly Gly Gly Ala Ile Thr Gly Met Gly Glu Ala Leu Thr Pro Ala Gly

			20					25					30		
Pro	Asp	Gly 35	Met	Ala	Ala	Leu	Ser 40	Leu	Pro	Leu	Pro	Ile 45	Ser	Ala	Gly
Arg	Gly 50	Tyr	Ala	Pro	Ser	Leu 55	Thr	Leu	Asn	Tyr	Asn 60	Ser	Gly	Thr	Gly
Asn 65	Ser	Pro	Phe	Gly	Leu 70	Gly	Trp	Asp	Cys	Asn 75	Val	Met	Thr	Ile	Arg 80
Arg	Arg	Thr	Ser	Thr 85	Gly	Val	Pro	Asn	Tyr 90	Asp	Glu	Thr	Asp	Thr 95	Phe
Leu	Gly	Pro	Glu 100	Gly	Glu	Val	Leu	Val 105	Val	Ala	Leu	Asn	Glu 110	Ala	Gly
Gln	Ala	Asp 115	Ile	Arg	Ser	Glu	Ser 120	Ser	Leu	Gln	Gly	Ile 125	Asn	Leu	Gly
Met	Thr 130	Phe	Thr	Val	Thr	Gly 135	Tyr	Arg	Ser	Arg	Leu 140	Glu	Ser	His	Phe
Ser 145	Arg	Leu	Glu	Tyr	Trp 150	Gln	Pro	Gln	Thr	Thr 155	Gly	Ala	Thr	Asp	Phe 160
Trp	Leu	Ile	Tyr	Ser 165	Pro	Asp	Gly	Gln	Ala 170	His	Leu	Leu	Gly	Lys 175	Asn
Pro	Gln	Ala	Arg 180	Ile	Ser	Asn	Pro	Leu 185	Asn	Val	Asn	Gln	Thr 190	Ala	Gln
Trp	Leu	Leu 195	Glu	Ala	Ser	Val	Ser 200	Ser	His	Gly	Glu	Gln 205	Ile	Tyr	Tyr
Gln	Tyr 210	Arg	Ala	Glu	Asp	Glu 215	Thr	Asp	Cys	Glu	Thr 220	Asp	Glu	Leu	Thr
Ala 225	His	Pro	Asn	Thr	Thr 230	Val	Gln	Arg	Tyr	Leu 235		Val	Val	His	Tyr 240
Gly	Asn	Leu	Thr	Ala 245	Ser	Glu	Val	Phe	Pro 250		Leu	Asn	Gly	Asp 255	Asp
Pro	Leu	Lys	Ser 260	Gly	Trp	Leu	Phe	Cys 265	Leu	Val	Phe	Asp	Туг 270	Gly	Glu
Arg	Lys	Asn 275	Ser	Leu	Ser	Glu	Met 280	Pro	Pro	Phe	Lys	Ala 285	Thr	Ser	Asn
Trp	Leu 290		Arg	Lys	Asp	Arg 295	Phe	Ser	Arg	Tyr	Glu 300		Gly	Phe	Ala
Leu 305		Thr	Arg	Arg	Leu 310		Arg	Gln	Ile	Leu 315		Phe	His	Arg	Leu 320
Gln	Thr	Leu	Ser	Gly 325		Ala	Lys	Gly	Asp 330		Glu	Pro	Ala	Leu 335	Val
Ser	Arg	Leu	11e 340		Asp	Tyr	Asp	Glu 345		Ala	Val	Val	Ser 350		Leu
Val	Ser	Val 355		Arg	Val	Gly	His 360		Gln	Asp	Gly	Thr 365		Ala	Val

Ala	Leu 370	Pro	Pro	Leu		Leu 375	Ala	Tyr	Gln	Pro	Phe 380	Glu	Pro	Glu	Gln
[3,7C*		Ten	Trp	Ara			Δen	Va1	Ten	Ala		Phe	Asn	Thr	Tle
385	Ala	beu	пр	ALG.	390	ricc	gen	var	Deu	395	LT:	EIIC	112.11	1111	400
Gln	Arg	Trp	Gln	Leu 405	Leu	Asp	Leu	Gln	Gly 410	Glu	Gly	Val	Pro	Gly 41 5	Ile
Leu	Tyr	Gln	Asp 420	Lys	Asn	Gly	Trp	Trp 425	Tyr	Arg	Ser	Ala	Gln 430	Arg	Gln
Thr	Gly	Glu 435	Glu	Met	Asn	Ala	Val 440	Thr	Trp	Gly	Lys	Met 445	Gln	Leu	Leu
Pro	Ile 4 50	Thr	Pro	Ala	Ile	Gln 455	Asp	Asn	Ala	Ser	Leu 460	Met	Asp	Ile	Asn
Gly 465	Asp	Gly	Gln	Leu	Asp 470	Trp	Val	Ile	Thr	Gly 475	Pro	Gly	Leu	Arg	Gly 480
Tyr	His	Ser	Gln	His 485	Pro	Asp	Gly	Ser	Trp 490	Thr	Arg	Phe	Thr	Pro 495	Leu
His	Ala	Leu	Pro 500	Ile	Glu	Tyr	Thr	His 505		Arg	Ala	Gln	Leu 510	Ala	Asp
Leu	Met	Gly 515	Ala	Gly	Leu	Ser	Asp 520	Leu	Val	Leu	Ile	Gly 525	Pro	Lys	Ser
Val	Arg 530		Tyr	Ala	Asn	Asn 535	Arg	Asp	Gly	Phe	Thr 540	Glu	Gly	Arg	Asp
Val 545		Gln	Ser	Gly	Gly 550	Ile	Thr	Leu	Pro	Leu 555		Gly	Ala	Asp	Ala 560
Arg	Lys	Leu	Val	Ala 565		Ser	Asp	Val	Leu 570		Ser	Gly	Gln	Ala 575	
Leu	Val	Glu	Val 580		Ala	Thr	Lys	Val 585		Cys	Trp	Pro	Asn 590		Gly
His	Gly	Arg 595	Phe	· Gly	Gln	Pro	11e		Let	Pro	Gly	Phe 605		Gln	ser
Ala	Ala 610		n Phe	Asn	Pro	Asp 615		(Va)	. His	Leu	Ala 620		Leu	Asp	Gly
Ser 625		Pro	Ala	a Asp	630		e Tyr	· Val	His	635		His	Leu	a Asp	640
Phe	e Ser	Asr	n Glu	Ser 645		Asr	ı Gly	/ Phe	Ala 650		n Pro	Phe	e Thr	655	
Phe	e Pro) Asp	Gly 660		ı Arç	Phe	e Asp	Asp 669	_	c Cys	s Glr	ı Let	Glr 670		. Ala
Ast	o Val	Glr 675	n Gly 5	/ Leu	ı Gly	/ Val	L Val 680		: Le	ı Ile	e Leu	. Sei 685	_	Pro	His
Met	Ala 690		o His	s His	Trp	Arg 699		s Asj) Lei	ı Thi	700		a Lys	s Pro	Trp
Le:		ı Ser	r Glu	ש Met	Asr 710		n Ası	n Met	c Gly	y Ala 71		s His	s Thu	c Lev	His 720



1065 1070 1060 Thr Pro Leu Lys Thr Pro Thr Arg Gln Ala Leu Ile Ala Phe Thr Glu 1080 Thr Ala Val Leu Thr Glu Ser Leu Leu Ser Ala Phe Asp Gly Gly Ile 1095 1100 Thr Pro Asp Glu Leu Pro Gly Leu Leu Thr Gln Ala Gly Tyr Gln Gln 1110 1115 Glu Pro Tyr Leu Phe Pro Leu Ser Gly Glu Asn Gln Val Trp Val Ala Arg Lys Gly Tyr Thr Asp Tyr Gly Thr Glu Val Gln Phe Trp Arg Pro 1145 Val Ala Gln Arg Asn Thr Gln Leu Thr Gly Lys Thr Thr Leu Lys Trp 1160 Asp Thr His Tyr Cys Val Ile Thr Gln Thr Gln Asp Ala Ala Gly Leu 1175 Thr Val Ser Ala Asn Tyr Asp Trp Arg Phe Leu Thr Pro Met Gln Leu 1190 1195 Thr Asp Ile Asn Asp Asn Val His Ile Ile Thr Leu Asp Ala Leu Gly 1205 1210 Arg Pro Val Thr Gln Arg Phe Trp Gly Ile Glu Asn Gly Val Ala Thr Gly Tyr Ser Ser Pro Glu Ala Lys Pro Phe Thr Pro Pro Val Asp Val Asn Ala Ala Ile Ala Leu Thr Gly Pro Leu Pro Val Ala Gln Cys Leu 1255 Val Tyr Ala Pro Asp Ser Trp Met Pro Leu Phe Gly Gln Glu Thr Phe 1270 1275 Asn Thr Leu Thr Gln Glu Glu Gln Lys Thr Leu Arg Asp Leu Arg Ile Ile Thr Glu Asp Trp Arg Ile Cys Ala Leu Ala Arg Arg Arg Trp Leu Gln Ser Gln Lys Ala Gly Thr Pro Leu Val Lys Leu Leu Thr Asn Ser 1320 1325 Ile Gly Leu Pro Pro His Asn Leu Met Leu Ala Thr Asp Arg Tyr Asp 1335 Arg Asp Ser Glu Gln Gln Ile Arg Gln Gln Val Ala Phe Ser Asp Gly 1355 1350 Phe Gly Arg Leu Leu Gln Ala Ala Val Arg His Glu Ala Gly Glu Ala 1370

Trp Gln Arg Asn Gln Asp Gly Ser Leu Val Thr Lys Met Glu Asp Thr 1385

Lys Thr Arg Trp Ala Ile Thr Gly Arg Thr Glu Tyr Asp Asn Lys Gly 1400

1405

Gln Ala Ile Arg Thr Tyr Gln Pro Tyr Phe Leu Asn Asp Trp Arg Tyr 1410 1415 1420	
Val Ser Asp Asp Ser Ala Arg Lys Glu Ala Tyr Ala Asp Thr His Ile 425 1430 1435 1440	
Tyr Asp Pro Ile Gly Arg Glu Ile Gln Val Ile Thr Ala Lys Gly Trp 1445 1450 1455	
Leu Arg Gln Asn Gln Tyr Phe Pro Trp Phe Thr Val Ser Glu Asp Glu 1460 1465 1470	
Asn Asp Leu Ser Ala Asp Ala Leu Val 1475 1480	
<210> 15 <211> 23 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 15 cgggatccga tgattttaaa agg	23
<210> 16 <211> 16 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 16 gcgccattga tttgag	16
<210> 17 <211> 19 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 17 cattagaggt cgaacgtac	19
<210> 18 <211> 26 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 18 gagcgagctc ttacttaatg gtgtag	26
<210> 19	

<211> 28 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 19 cagcgagctc catgcagaat tcacagac	28
<210> 20 <211> 18 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 20 ggcaatggca gcgataag	18
<210> 21 <211> 18 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 21 cattaacgca ggaagagc	18
<210> 22 <211> 26 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 22 gacctcgagt tacacgagcg cgtcag	26

```
SEQUENCE LISTENS
<110> Novactie AG
<120> Nove), Toxins And Uses Thereof
<130> PI/5-30421/A/CGC 1963
<140>
<141>
<160-22
<170> PatentIn Ver. 2.0
<210> 1
<211> 9717
<212> NA
<213> Photorhabdus luminescens
<220>
<221> CDS
<222> {412}..(1665)
<223> arf1 ~16.4 kbs
<220>
<221> cms
<222> (1686) ... (2447)
<223> orf2 ~28.1kDa
<220×
<221> CDS
<2225 {2750}..(3310)
<223> orf3 ~20.7 kpa
<2205
<221> CDS
<222> (3342)..(4118)
<223> ort4 -28.7 kDa
<220>
<221> CDS
<222> (4515) . . (9269)
<223> orf5 -176 kDa
gsattcatat gotatgaeat asacegitigg cycaataett asacciatta tittiatiit 60
gttüttätas aatgatatgo tittattaass agaataatga gttästgota aataaatoot 120
oggstitet cetgetetta iggccgesig igstgigesc valtatitta taattagati 180
aztaslálas figgiattasa atascestal etitalicet iggitattat catoggitti 240
attacatggy gaalaatola taaattagti tiacataatt cacaaatagc qattacatta 300
accoggaata trassasata: trattratga rhatgatgat ababetreat ragectaert 360
ttataactag esaaattgac attiticaate catgitateaa tgytaeccaa t atg ceg 417
                                                            Met Gln
```

age got case cos got stt got ace gog att Arg Ala Gln Arg Val Vel Ile Thr Gly Met Gly Ala Val Thr Pro Ile 5

ggr. Gly	Glu 20	gat Asp	gtt Val	gaa Glu	Ser Ser	tgt Cys 25	tgg Tvp	caa Gln	agt Ser	att Ile	att Ilo 30	gaa Glu	833 (2)/8	caa Gin	cat. His	513
eça Arg 35	ttt Phe	cac His	aga Arg	att Ile	വരു Glu 40	ttt Phe	cet Pro	gold Asp	tca Ser	tte Phe 45	att Ile	aat Ayrı	teg Ser	egt. Arg	tte The 50	561
ttt Phe	tet Ser	tto Ehe	ctt Leu	gra Ala 55	cca Pro	aac Aso	cca Pro	tor Ser	ege Arg 60	tat Ty r	ceg Gln	tta læn	tta (x)u	eca Pro 65	aaa lys	609
аэд Јус	ttg Jen	act Thr	cat Hie 70	oca Thr	ctt Leu	tct Ser	ved Dac	bge Cys 75	gga Gly	een Lys	gca Ale	gca Ala	ttg Leu 80	TAS Sed	gog Ala	657
Thr	Тут	GLn B5	Ala	Fire	Thr	Glm	Ala 90	tte Phe	Gly	Val	Aen	I.) e. 95	දින	Sto	Val	705
Glu	100 100	Тут	Asp	рÀs	<u> ጌ</u> አን:	Glu 3 0 5	Cya	gj ^A 38c	Val	lle	Leu 110	Cly	Ser	Ġĺy	Trp	753
Gly 115	Мà	IJB	Asp	Asn	Ala 120	GlY	Азр	ræt His	Ala	Cys 125	gln	Тут	Lyrs	Gla	Ala 130	601
Lys	Lou	Ala	Hás	ਜ਼ਿਨ 135	Mat	Sar	Asn	ctt Læ	110 140	Tru	Met	Pro	Ser	<u>Ser</u> 145	Met	849
Thr	Ala	Alā	Cys 150	Ser	Ile	Met	Tyr	gga Gly 155	تجها	Arg	Gly	ፒሃፕ	Gla 160	aen Cest	Tîur	B97
Val	Met	Ala 165	Ala	Cys	Ala	Tir	Gly 170	aca Thr	M≘t	Ala	Ile	©Ту 175	A.∓D	Āla	Fhe	945
Glu	11e 180	He	Azg	Sour	Gly	Arg 185	Ala	ess Lys	CAa	M≘t	11e 190	АÌà	Gly	Ala	Āla	993
GIM 195	Ser	ſ₩	Thr	Axg	31u 200	్రజ	Asn	att Ile	Ттр	පිත 205	Пе	Аєр	Val	Len	Asn 210	1041
Ala	Lou	Sor	Lys	Glu 215	Gln	Ala	Авр	cca Fto	AB D 220	Leu	Ala	Cy#s	Cys	Pro 225	בתיו	1089
age Ser	citti Leu	gal. Asp	ogo Arg 230	tet Ser	GJA 339	ttt Phe	gta Val	tta Leu 235	goc Ala	gaa Glu	суу Сору	gog Ala	903 Ala 240	gua Val	gtit Val	1137
tgt Cys	ctg Leu	922 Glu 245	aat Asn	tai. Tyr	get: Asp	tca Ser	gu: Ala 250	eka Ile	Ala	Arg Cgt	gjy Gjy	gca Ala 255	acg The	att Ile	tta Lau	1185
Me Als	988 Glu 260	att. Ile	aca Lye	ggt Gly	tac Tyr	ясс Ala 265	caa Gln	tat Tyr	tca Ser	gat gu4	gee Ala 270	yal Yal	aa(; Asn	ita Leu	&CC Thr	1233

ogg Arg 275	rce Pro	aca Tha	gaa Gî∪	get Aug	att Ile 2B0	ருர் வர	ect Pro	aaa Lys	ata Ile	tta Leu 285	YJ# DOI	ata I.le	act Thr	aaa Lys	god Ala 290	1281
att Ilc	()) n මෙර්	Gin cag	gca Ale	cag Gln 295	att Ile	tog Ser	coş Pro	aaa Lye	gat Asp 300	att Ile	yeşd Asşd	tac T y r	att Zje	aat Asn 305	get Ala	1329
cat His	ggt. Gly	açit Tim	tet Ser 310	aca Thr	ocg Pro	tts Leu	aat Asn	gat Asp 315	ctt Leu	tat. Tyr	gna Glu	act Thr	തു Glin 320	gca Ala	att Ile	1377
əəa Lye	gce Ala	gça Ala 325	<i>c</i> tg Læu	gge Gly	CBB CBB	tet Tyr	get Ala 330	tat Tyr	ලාව සෙබ	gte Val	eet Pro	ata Ile 335	tox Ser	agc Sec	ace Thr	1425
F9/8 Essa	tet Ser 340	tat Tyr	acc Thir	CIY GGC	cac Kis	ctt Leu 345	act Ile	get Ala	ger Ala	gec Ala	99t. Giy 350	agt Ser	ttt Phe	gaa Giy	acg Tha	1473
att Ile 355	gta Vəl	tet Cya	gtg Val	ana Lys	90a Ala 360	tta Leu	gct Ala	gaa Glu	aat Asn	tge Cys 365	ttg Leu	eca Pro	gca Ala	sco Tur	ttg Leu 370	1521
aat Aen	tta Leu	cac His	Arg Ogg	906 Ala 375	gat Asp	oca Pro	gat Ago	tgc Cys	gat Aep 380	ete Leu	aat Asn	tat Tyr	ttg Leu	oct Pro 385	eat Aen	1569
cza Gln	cat His	tgc Cys	tac Tyr 390	acc Thr	Ala Sct	can Gln	oca Pro	დაც Glu 395	gtg Vel	aca Thr	ctc Leu	eat Asn	att Ile 400	Agc Ser	gca Ala	1617
gly ggt	tte Pha	ддс (31 _У 405	01 ⁷ . 88 в	cat His	yzu 8ac	get Alæ	geg Ala 410	ttg Leu	gtt Val	atc Ile	get Ala	аад Гус 415	gta Vel	agg Atg	taa	1665
ctga	itatç	itt (jatti	Etgi	nsat He	et G	aa ge lu Aa 20	at ad ap 11	it gs le Gi	ea ce lu Hi	is T	ညာ လ ကျော့ ဆ 25	я. У. ЭЗ 99	at ti Sci Pi	ic tet 16 Ser	1716
999 Gly 430	gat Æp	ttt Phe	aac Asn	occ Pro	ato Ile 435	cat His	tat Tyr	teg Ser	geg Ala	298 Lys 440	ayr: Ser	G)u gag	tot Ser	ttg Lev	ogc Ary 445	1766
aat Asn	ata Ile	cag Gln	caa Gln	cac His 450	eeg Pro	gtg Val.	Cag Cag	gga G)y	atg Met 455	ttg Leu	egt Ser	tte Leu	ete Leu	tat 3yr 460	gta Vøl	1814
arg Arg	cea Gln	cag Glin	ttt Phe 465	tot Ser	Gln Can	tta Leu	art Thr	toc S≥r 470	get Als	ttt Rbe	aca Thr	Jpr. Ord	9 <u>9</u> व्ह द्वी <i>पू</i> 475	ata Ile	ttg Læi	1862
aac Asn	att Ile	gat Asp 480	gcc Ala	tet Ser	tte Pha	ege Arg	сад 61л 485	tat Tyr	gtt Val	tat Tyr	ecc Thr	900a Ala 490	tta Leo	ecc Pro	cat Kis	1910
Gln Gln	ctg Lev 495	agg Avg	att Ile	aat <i>As</i> n	act Thr	өг 1978 500	aac Aen	ryz ees	acy Thr	ti:t Phe	aaa Lye 505	tta Leu	gaa Gin	aat Axan	ecc Pro	1958
agt Sor Sio	aaa Lys	gna Cilu	aac එ ණ	aog Thr	ttg Leu 515	ttc Phe	GIY ggc	eet Asn	jju ecc	age Ser 520	gta Val	gag Glu	aat Asn	aca Thu	atg Met 525	2006
									gat							

Glv	Ser	Ile	Glu	Авр 530	Trp	Ile	Val	Gln	А гр 535	ሕርብ	Cys	Cln	lye	Leu 540	Thr	
ata Ile	aca Thr	eg?. eata	929 Glu 545	Glu Glu	gtt Vel	tgt Cys	gaa Glu	ээд Цув 550	tat Tyr	get Ala	gtc Val	t.t.l: Phe	aga Arg 555	tac Tyr	tat Tyr	2102
ttc Pre	oca Pro	agt Ser 560	gtc Val	act Thr	tet Ser	att Ile	992 Gly 565	tgg Trp	tte Phe	ctg Leu	gat Asp	gog Ala S 7 0	ctt Leu	gct Ala	ttt Phe	2150
cat Me	otb Len 5 75	att Ile	att Ile	aat Aen	tog Ser	aca Thr 580	gga Gly	ttt Phe	ett Læj	aat Ası	ttt Phe 585	GJπ G≊G	C&C His	tac Tyr	cat Kis	2198
ttt Ehe 590	asc Aen	coa Gln	tta Leu	Glu cab	gat Asp 595	tat Tyr	cts Leu	agt Ser	caa Gln	tet Sen 600	ttt Fhe	act Thr	t:tg Leu	cat His	act The 605	224€
61y 999	<u>СЭЭ</u> СЭЭ	gog Ala	att Lle	а а а Lys 610	atc Ilo	agg Arg	පනය දැව	GJ r Gæg	att Ile 615	gtt Val	eat Aen	agt. Ser;	aca Thr	gta Val 620	tta Leu	22 9 4
tta leu	tet Ser	toa Ser	00g Pro 625	gat Asp	ate Ilo	tgt Cys	gtt Val	gaz Glu 630	tte Leu	aat Aen	oct Pro	cct Pro	tta Lou 635	ttg Leu	att. Ile	2342
aag Lys	aat Aen	gạc Gly 640	gent Aep	aaa Lys	gat Asp	tat Tyr	att Ile 645	egt Arg	att Ile	tte Rhe	tat Tyr	tat Tyr 650	cga Arg	tgt. Cys	tita Leu	2390
tet Tyr	gat. Asp 655	ana Lys	aga Lys	cet. Pro	att Ile	ttt Phe 660	gta Val	tca Ser	rag Lys	act The	tes Ser 665	att 11e	atc Ile	tot Ser	aag Lys	2436
atg Met 670	aaa Lye	taa	aagg	jaaa j	gcy a	seetg	yodbe	uc ac	28.000g	ates:	att	ttc	etg			2487
ooat	aaa,	yaa t	agaa	atatt	ta at	gato	ووحجز	g ate	teg	intera	tgaz	gaas	ata s	bCB(1	agagt.	2547
oote	tttt	gt t	togo	ttga	a tt	.tgat	agto	e ttg	æcte	atet	950%	ato	aa g	tct.	lytgt	2607
togge	M O OS	nta b	ggta	ttgt	got	taaa	geeg	; aac	ttt	ttc	apot	catt	ct a	t to tar	200a t	2667
taas	tgaş	et c	acto	pacte	at tt	аааа	tcaa	aat	tgte	wate	tgaz	ittt	ta c	ttee	shtatg	2 727
ttit	itt er	icc a	itta:	catt	a aç	aggt	tate	atg y et	aec Aer	: gt(: Va) 675	Lea	ı çası ı Glu	Car Clr	ggt Gly	. 689 11ye 680	2781
gtt Vel	gct Als	get Als	tta Leu	tat Tyr 685	tes Ser	VJS ÖCC	Let: Tyr	tog Ser	690 Сјл Эаз	aca Thr	ÇESE Gl.u	ggt Gly	tet. Ser	tog Ser 695	tgg Trp	2829
gtg Val	goz Gly	aac Asn	ttg Leu 700	tge Cys	tgt, Cys	utt. She	toa Ser	agt Ser 705	gat Aep	yr.d. côd	gag Glu	cat Kis	tilg Leu 710	cct Pno	att Ile	2677
ato Ile	gtg Val	aat Asn 715	GJA ååå	egt Arg	cgt Arg	Phe	ttg Leu 720	att Ile	902 902	ttt Pie	gtt Val	att Ile 725	cca Pro	gat Aвр	cat Hi≋	2925
tta Leu	ct.t. Leu	gest. Carres	aaa Lys	acg Thr	gtt Val	aaa Ly⁄≥	CCC Prt)	эда Эда	gta Val	ttc Pha	gat Asp	ttg Leu	gat Nego	atc Flo	aat Asn	2973

- 5 -

730	735	740
ama cam tit the chy cyt lys Glm Phe Leu Leu Arg 745 750	ogt gae oat oat gag Arg Asp His Arg Glu 755	ata aat att tet oft - 3021 The Asm lle Tyr Leu 780
the got gee ggs eat thi. Leu Gly Glu Gly Asn The 765	atg got ogg sog ecg Met Asp Arg Thr Thr 770	aca gat eas ent ota - 3069 Tur Asp Lys Asm Leu 775
tto gay tta aat gag gat Phe Glu Leu Asn Glu Asp 780		
oct ctt ogt aan tat gtt Ala len Gly Lye Tyr Val 795	get att aat ees tee Ala Ile Asn Pro Ser 800	act acg cas tit atc 33,65 Thr Thr Oln Phe 11e 805
tte ttt gea ega gga sag gng Phe Ala Gin Gly Lye 810		
aca git gaa gar gaa tix Thr Val Glu Amp Glu Leu 825 830	Ser Lys Arg Tyr, Arg	gtc aga att att oot 3261 Vnl Ary Ile Ile Pro 840
gaa tig maa ggy (cog bat Glu leu Gln Gly Pro Tyr 845	tat gge tit gas eit Tyr Gly Pre Glu Leu 850	gal att ett tet att 3309 Amp Ile Leu Ser Ile 855
aca got taa ttoacaatat Tho Ala		aa eeg een ato oon aca 3362 lu Lya Lya Ile Thr Thr 865
tit atc att gag ama mot The Thr Ile Glu Lys Thr 870	gat gad aat til; tot Asp Asp Asn Fhe Tyr 875	not aat ggg ogt cas. 3410 Als Ass. Cily Arg His 680
cam tyt mty ytm mmn ato Gln Cym Met Val Lys Ile 885		gaa tat agg Ast 99t. 3458 Glu Tyr Arg Amo Gly 895
Gln Cys Met Val Lys Ile	: Acor Val Lau Dys Glm 890 a coot ago gag got gwa	(Glu Tyr Arm, Aen Gly 895 (aea agas teg att cag 3506
Gln Cys Met Val Lys Ile 865 gat tgg ats aas Cto goo Asp Trp Ile Lys Leu Ala	s Sor Val Lau Lys Gli 890 a cit agt gag gct gas s Leu Ser Glu Ala Glu 905 ago cto ata tat gas o Ser Leu Ile Tyr Asg	(Glu Tyr Arm Asn Gly 895 (ass angs try att cag 3506 (lys Arm Ser Jie Glu 910 (cas the ass atg cet 3554 (Clu Leu Lys Met Pro
Gin Cys Met Val Lys Ile 865 gat tgg ats aas Cts goo Asp Trp Ile Lys Leu Ala 900 gtg gog goa tta agt gat Val Ala Als Leo Ser Asp	s Sor Val Lau Lys Glu 890 a ott agt gag got gas s Leu Ser Glu Ala Glu 905 a ser Leu Ile Tyr Asy 925 a gal goa aga aat aas	assa nga teg att esg 3506 lassa nga teg att esg 3506 lays Arg Ser 12c Gln 910 cras the mea atg cet 3554 coll Leu Lys Met Pro 930 attt gat ett ogg tta 3602
Gln Cys Met Val Lys Ile 865 gat tgg ats aas Cta got Asp Trp Ile Lys Leu Ala 900 gtg gog goa tta agt gat Val Ala Als Leu Ser Asp 915 too ggt tgg aca acg acg Ser Gly Trp Thr Thr Thr	s Sor Val Lau Lys Glu 890 a ctt agt gag get gas s Leu Ser Glu Ala Glu 905 a ago ctc ata tat gas o Ser Leu Ile Tyr Asp 925 a gal goa aga aat aas r Asp Ala Ang Asu Lys 940 t get gat get ttt al.	Solution and Arm Aem Gly 835 also age tog att cag 3506 lys Arg Ser Ilo Glm 910 cas the age atg cet 3554 Cln Leu Lys Met Pro 930 attt gat ctt ggg tta 3602 a Phe Aep Leu Cly Leu 945 cas gac gas cag gte axe 3650
Gln Cys Met Val Lys Ile 665 gat tgg ats aas Cta got Asp Trp Ile Lys Leu Als 900 gtg gog gos tta agt gat Val Ala Als Len Ser Asp 915 tos ggt tgg acs acg acg Ser Gly Trp Thr Thr Thr 935 tta aat ggt gtt tal cat low Asn Gly Val Tyr Kir	s Sor Val Lau Lys Gluss 890 a cit agt gag get gas leu Ser Glu Ala Glu 905 age cit ata tat gac ser Lau Ile Tyr Asp 925 a gat gea aga aat aas r Asp Ala Arg Asn Lys 940 t get gat get tit All s Ala Asp Ala Phe Ile 955	assa age tog att cag 3506 lys Arg Ser Ilo Gln 910 cras the cea etg cet 3554 collin Leu Lys Met Pro 930 de tit gat cit ggg tia 3602 de Phe Asp Leu Cly Leu 945 c gac gma cag gia ace 3650 de Asp Glu Gln Val Thr 960 de tial cag ace agt gig 3698

age ara gam two eta aty sea am aty aen tit gam gat mey gat gag Ser Thr Glo Tyr Leu Met Alm Lys Met Thr Phe Glo Amp Thr Amp Gly 995 1000 1005 1010	3794
aas ogo aca tta eca eog aat atg toa gtt ggt gat gew gtt tit gad lys Arg Thr Jev Tho Tho Ash Mel. Sen Val Cly Asp Glu Val Phe Asp 1015 1020 1025	3842
age aag git tie tie aan goe aft got oot tat gee aft aal aee aaf Ser Lye Val Leu Leu Lys Ala Ile Ala Pro Tyr Ala Ile Asm Thr Asm 1030)035 1040	3890
one tig cat see and aid and are tig tit gut amm aca gen geg cog Gln led His Glu Aen Ile Aen Thr led Phe Asp Lys Thr Glu Glu Pro 1045 1050 1055	393B
aca and too got not that out one ath aft and off that one that aca Thir lys Ser Asp Thir Mis His Glin He He Ash Leu Tyr Ang Thir 1060 1065 1070	39B6
the cos tat can the age ath chi. San egg not see agt act ett eat Leu Pro Tyr His Leu Arg lle Leu Glu Gly Asn Asp Set Thr Val Asn 1875 - 1880 - 1885 - 1889	4034
aga ata tat gic cit ggt asa gag cca toa eau got aga tic cig ace Arg Ile Tyr Vel Leu Gly Les Clu Pro Ser Asn Asp Arg Pho Leu Thr 1895 1180 1185	4082
aga gga agg gta tit aaa oga gga act cat aig iga aigracgiga Arg Gly Arg Val Phe Lys Ary Gly Thr His Met 1110 1115	4128
tastgigagi ggaggeigig thabaysobs textiataco gtaactaitc oggacxopca	4188
perference gaagegette augegaeagg gegetrojegg cognagegytt autabgatog	4248
atabratgat gicacaatca tigataaria rggrigboag cataaattia gaattictic	430B
ggittaatait ggaogigogo taagrafago: gagaxtaagi igaittiicol: bagitamaaa	436B
cettigitta igetggiaaa ogealgigeg likgomagom allaatalat Uxmatatig	442B
aastaggaat atageestat etgtasttat acetasaega attittaete yaatataatt	448 8
ttantigate assenggasa ittaas eig aan get ace get ata tat iku aat. Mei Lys Ala Thr Asp Ile Tyr Ser Am 1120 1125	4541
got tit ami tito ggt tot tal sit ami ack ggt gto gmi occ aga akw Ala Phe Aen Phe Gly Sar Tyr lle Aen Thr Gly Val Aep Pro Arg Thr 1130 1135 1140	4 589
ggt caa tat agt gca ast all, est elt atc acc tta aga cot ast act Oly GL) Tyr Scr Ala Asn Ile Asn Ile Ile Thr Lea Arm Pro Asn Asn 1145 1150 1155	4637
gtg ggt aat toy gaa can aco ttg ago ota tos tto toy oua tta aca	
Val Gly Asa Ser Glu Gla Thr Leu Ser Leu Ser Pho Ser Pro Leu Thr 1160 1165 1170 1175	46 B5

tta gat ata was ace cit aca tit age oga gea aat gag ga Leu Aep Ile Iya Thr Leu Thr Pha Sor Azy Ala Ash Gly Gl 1195 1200	u Gln Ihe
ann tot ang own ilig ong ont ant ant ant gat oft agt if Lye Cye Lye Pro Leu Pro Pro Asn Asn Asn Asp Leu Ser if 1210 1220	
aas aan ota aas gat tig oge gis tat aag ote gat age as bys Lys Lou Lys Asp Leu Arg Val Tyr bys Leu Asp Ser Ag 1235 1235	
thi git thi and man man gigo att also gay ata cit man or Tyr Vol Tyr Asm Lys Asm Gly Ile Ile Glu Ile Leu Lys As 1240 1250	p att 990 4925 9 Ile 91y 1255
tog agt gat att goa aam acm got gom ott gam tit oot ga Sor Sor Asp The Ale Lys Thr Val Ale Leu Glu Phe Pho As 1260 1265	atggrtgsa 4973 ap Gly Glu 1270
gra tit gat ita att tat aat tea aga tit gea tig tee ga Ala fhe Asp Leu Ile Tyr Asm Ser Arg fhe Ala Leu Ser Gl 1275 1280 128	w Ile Lyc
tec cot ots see got as a est tat stt as stee as tec to Tyr Arg Val Thr Gly Lys Thr Tyr Leu Lys Leu Asm Tyr Se 1290 1300	ot gga aat 5069 er Gly Asn
Asn Cys Thr Ser Val Glu Tyr Pro Asp Asn Asn Asn Ile Se 1305 1310 1315	et gog ama 5117 er Ala Dys
ata goa tto gat tat ogt aac gat tac ott att acg gtg ac Ile Ala Phe Asp Tyr Arg Asm Asp Tyr Leu Ile Thr Val Th 1320 1325 1330	rt gta cet 5165 or Val Pro 1335
tad get get tet get oot att get tet gee oge tit mæg af Tyr Asp Ala Ser Gly Pro Ile Asp Ser Ala Arg Phe Lys Me 1340 1345	tg acc tat 5213 el: 15t/ Tyx 1350
cag acs tha saa ggc ghs thi com ght she age acc the or Ghn Thr Law Lys Ghy Val Phe Pro Val The Ser Thr Phe 20 1355 1360 136	rg Thir Pro
acc got tat get gag rtg gtg agt tat aaa gag aat ggg ca Thr Gly Tyr Val Glo Leu Val Scr Tyr Lys Glo Asn Gly H 1370 1380	at aan gtg 5309 is Lys Val
any gan any gam tat att oot tat gog got gos ote act at Thur Asp Thur Glu Tyr Ile Pro Tyr Ala Ala Ala Leu Thu Il 1385 1390 1395	
ggr aat gga caa oot gng gto ago aaa too tat gga tat ag Gly Aso Gly Glo Pro Ala Val Ser Lys Ser Tyr Glo Tyr Se 1400 1405 1410	gt toa gta - 5405 er Ser Val 1415
cat aac the the gen hat but but one one can ago the on His Asm Pho Leu Gly Tyr Ser Ser Gly Arm Thr Ser Phe Ar 1420 1425	at toe agt 5453 sp Ser Ser 1430
caa gat aat tig tat tig gid aca ggg ama tad sot kat to Glo Asp Ash Leu Tyr Leu Val Thr Gly Lys Tyr Thr Tyr So 1435 1440 144	er Ser Ike
gas ogg git its get ggt caa agt gig git ich gia ata gr	aa oga gta 5549

GIU		VBI 145D		Asp	GIA		Ser 1455		Val	Ser		Ile 1 4 60		A) T	Wal	
Pha	aat As n 1 66 5	Lar:	ttc Phe	cat Nie	Leu	atg Met 1470	ljur eroc	aaa Lys	GJU GBB	Ala	886 Lye 1475	aca Thr	cáit Glin	gat Asp	aat Aen	5597
ಕಿತವು 199 1480	Arg	att Ile	ace Thir	Thu	gee Glu 1485	att Ile	act Thr	tec Tyr	aat Aso	gag Glu 1490	gat. Asp	cra Leu	tca Ser	Liys	agt Ser 14 9 5	5645
tto Phe	bça Ser	gag Glu	a	сса 2то 1500	gaa பெ	aat Asn	tta	Cin	caa Gla 1505	ect Pro	tet Sor	ege Arg	Val	tta Leu 1510	ecc: Thir	5693
æt Arq	tat Tyr	Thr	gat <i>As</i> p 1515	ata Ile	caa Gln	aca Thu	A(z)	act Thr 1520	tca Ser	oga Arg	gaa Glu	Glu	act 11r 1525	Val	ast Agn	5743
att Iìc	Lyx	əgt Şer 1530	yeb	gat Asp	tgg Trp	Gly	ааt Авп 1535	act Thr	eta Leu	ctt L⇔,	IJœ	act Thr 1540	gag Glu	acc Thr	agt Ser	5789
GIÀ	ata Ile 15 45	GJV CSÖ	aae Lys	gae Glu	Tyr	gtf. Val 1550	tat. Tyr	tat Tyr	rog Pro	Val	aat Aen 1555	ely gg:	gaa Glu	ej A ûût	aat Jen	5837
agt Ser 1561	CA≊	oct Pro	goc Ala	Vab	ece Pro 1565	ttg Leu	ggt Gly	ttt Yhe	tet Ser	099 Arg 1570	tte Phe	tta Leu	aaa Lys	Ser	gtt Val 1575	5885
acy The	Gln cea	TA:2 See	GLY	teg Sør 1580	eet Pro	get Asp	get Ala	Alz	മു Gln 1585	agt Ser	gtc Val	gon Ala	Aei y	а <u>аа</u> Lye 1590	AFJ Öfð	5 93 3
att Iie	cat Rie	Tyr	aca. Thir 1595	tat Tyr	caa Gln	aaa Lye	Phe	eet Pra 1600	act Thr	ttt She	acc The	Gly	get Ala 1605	tat Tyr	gti: Val	5981
ಗೌನ ತಾಡಿ	GIII	tat Tyr 1610	gtc Val	agt Ser	aaa liys	Val	tca Ser 1615	Gjn ðsið	acg The	ata Jl⊕	Asp	aat Asn 1620	ees Lys	eta ile	gog Ala	6029
Arg	acc Thr 1625	ttt Phe	ggo Ser	tat Tyr	ver	aac Asn 1630	toa Ser	cog Pro	ecg Thr	SAT	ава Lys 1635	tet Sar	cet His	ggt Gly	teg Ser	6073
tta Leu 1.6 4 0	MA	Lys Lys	eta Ile	Thr	toa Ser 1645	gtg Val	atg Met	eat Azn	ABD	cag Gln L650	can Gln	acg Thr	gtc Lev	Thr	aca Thin 1655	6125
ttt Bhe	lye Lye	tat Tyr	GIU	tat 'Tyr L660	tçe Ser	gaa Glu	agt Ser	Glu	ate Met 1665	ace Thr	aca Tho	aat Aan	Ala	acg Thr 1.670	gtg Val	6173
acc Thr	ggt Gly	Pto≘	gat Aep 1675	gg:	Ala Sca	cat Hi≱	7.44 C	ç sa Glu L680	teg Sor	aaa Lys	aat. Aso	Val	acy Thr 685	tet Ser	att Ile	6221
tat Tyr	TIT	cat His 169D	cgg Arg	caa Gln	ett Leu	Arg	ааа Lys 1695	gtt Val	gat Asp	gta Vel	শ্ৰুমা	cac Ris 1700	gtg Val	att Ilø	acc Thr	6269
gat Asp	cag Gln	tot Sen	tet Tyr	gst Asp	ctt Leu	ttg Leu	ggt Gly	ogo Ang	att Ile	aca Thr	999 Glv	cza Gln	att lle	att Ilo	gst Aso	6317

1705		Э.	71 D		1715			
၀၀င ကာင (Pro Gly 1 1720	aog gca Thr Ala.	ega gaa (Arg Glu) 1.725	att asa lle bys	Arg Aso	tac gtt Tyr Val 1730	tat caa Tyr Gln	tel: ccc Tyr Pro 1735	6365
Oly Oly i Oge pat (esto Cyn	aat gel (Aen Aep (740	titi teg Phe Trp	ceg gtg Pro Val 3745	atg ata Met Ile	Glu Val	gat tet Asp Ser .750	6413
caa gge ! Gln Gly !	ptc aga Val Arg 1.755	cyt ese a	Itor His	tac gat Tyx Asp 760	gga elg Gly Met	ggs: cgt Gly Arg 1765	att tgt Ile Cye	6461
tog att (Ser Ile (1	gee gee Glu Glu 770	cea get (Gln Asp /	gat gat Asp Asp 1775	gge gee Gly Ala	Typ Gly	aca tog 'Nor Ser 1760	ggg att Gly Ile	6509
tat cea q Tyr Gln (1785	ggreca Gly Thr	Tyr Arg I	eaa gtt Las Val 790	ctt gec Iæu Alæ	aga can Arg Gln 1795	tat ggt Tyr Aap	gtt tto Val Leo	6557
ggg ceg : Gly Gln : 1800	ttg age Læy Ser	аво два ; Lys Glu ; 1805	att toa Ile Ser	Aen Asp	tog tta Trp Lev 1810	tgg aat Trp Aso	tka tet Lau Ser 1815	6605
groc aat (Ala Asm)	Pro Leu	gtt ogt (Val Arg) 820	ett get Leu Ala	axx cxy Thr Pro 1825	tbg gtt Læu Val	Thr Thr	aga acc Lys Thr 1830	665 3
tat ana (Tyr Lys '	tat gat Tyr Asp 1635	gat tag (Gly Trp (Cjy yeu	ott tan Leu Syr .840	age acg Ser Thr	gaa tac Glu Tyr 1845	agt gat Ser Asp	6701
ggt ogg (Oly Arg) 1:	ata qag Ile Glu 850	etg gaa. Leu Glu	ato cat 1,1e Mis 1855	gat ect Asp Pro	Ile Tar	ayg aw Arg 11ur 1860	att act Ile Thr	6745
cas 999 (Gln Gly 1 1,865	gic eas Va l lys	Gly Leu	999 atg Gly M et 870	1.12a ast. Leu Asn	alt cog Ile Gln 1875	csa aat Gln Asn	aat ttt Aso Phe	679)
gag caa (Glu Glu 1880	oog get Pro Ala	tog atc Ser Ile 1 1885	aaa get Lys Ala	Val Tyr	oot gat Pro Asp 1890	ggt acg Gly Thr	ata tet 11e Tyr 1895	5845
age acc Ser Thr i	Arg Thr	tat cgt Tyr Arg ' .900	tat gat Tyr Aep	gga ttt Gly Phe 1905	CJA yrd GET CET	That Val	acy gas Thr Glu 1910	6 89 3
aca gát (Thr Asp .	gca gwa Ala Glu 1915	ggt cat Gly Hie .	Ale Thr	caa alt Gln Ile 1920	99a tat Gly Tyr	gat gtg Asp Val 1925	ttt get Phe Asp	6941
cgt ata (Arg Ile 1	gtg ame Val lys 930	aaa acg Lys Thr	ttg cca Leu Pro 1935	yad gya Ger Gös	Thr Ile	tta gea Leu Glu 1940	toe get Ser Ala	6989
tat gca Tyr Ala 1945	age ttt Ser Phe	Ser His	gaa gaa Glu Glu 950	tta att Lou Ilm	tog goe Sor Ala 1955	Leu Asn	gtg ælt Val Agn	7037
ggo aca Gly Thr (1960	cag ttg Gln Low	gyg gca Oly Ala 1965	ttagtt Lou Val	тус жер	ggt ett Gly le u 1970	Gly Arg Gyy Cgg	gta ata Val Ile 1975	7085

agt gal 40 Ser Asp Th	g gtg ggt ir Val Gly 1980	ggt oge a Gly Arg L	aa acg qaa ys Thr Glu 1985	Tyr Lea Tyr	ggg eet caa Gly Pro Gln 1990	7133
ogt sec as Gly Asp ly	a ong att s ivo lle 1995	cag tca a Gln Ser I	tt act eet le The Pro 2000	ing cat aat Ser Ris Asu	aag caa aat Lys Gin Asn 2005	7381,
atg get ta Met Asp Ty 201	a len Lat	tat off g Tyr Len G 20:	ly Ser Wal	atg toe eas Mot Ser lys 2020	ttt acc acg She Thr Thr	7229
999 ace 96 Gly Thr As 2025	p Gln Gln c cas caa	een tht d Aen Phe A 2030	9t hat cat ng Tyr His	tog aaa acg : Ser Lys Thr 2035	ggs aca tta Gly Thr Leu	7277
tta tet go Leu Ser Al 2040	a Ser Glu	990 900 U Gly Val S 2 04 5	et cag act er Gln Thr	aat tac agt Aan Tyr Sor 2050	tat tte cea Tyr Pha Pro 2055	7325
tog ggt g1 Ser Gly Va	a tta cag 1 Leu Gln 2060	rgs 908 () Arg Glu S	es til tta er Phe Lev 2065	i Ang Aspo Asn	aaa cog att Tau Pro Ilm 2070	7373
toa tog go Ser Ser Gl	r gag tar y Clu Tyr 2075	ctt tat a Leu Tyr T	og atg ted or Mat Ser 2080	Gly Leu Ile	can ogt cat Gln Arg His 2085	7421
a 28 gat ag Lya Asp So 209	r Pac Oly	cat ast c His Asm H 20:	is Val Tyr	agt tac got Ser Tyr Asp 2100	get eag gga Ma Gin Gly	7469
aga ttg gt Arg Leu Va 2105	c maa loa l Lye Thr	gaa cag g Glu Gin A 2110	at gos caa ap Ala Cln	ter got aca Tyr Ala Thr 2115	ttt gan tat Phe Glu Tyr	7517
gac sat gt Asp Asn Vo 2120	d Gly Arg	tte ata a Leo Ile T 2125	on alog alog ar Thar Tha	acc ama gac The lys Asp 2130	acy acy tra Thr Thr Ser 2135	7565
tta ter es Leu Ser Gl	e tia glg n Leu Val 2140	aca ase a The lys I.	to gava tat le Glu Tyr 2145	'Asp Ala line	gat oga gaa Aap Arg Glu 2150	7613
ata aaa og Tie Lys ar	e tog eta g San Leu 2155	ett agt ga Ile Ser As	o tto toa Op Phe Ser 2160	The Gln Val	att acc tta The Thr Leu 2165	7661
age tat ec Sex Tyr Th 217	r Lys Aso	eat caa a Asn Gln II 211	ie Ser Ch	egt ato acc Arg Ile Thr 2180	tee ate gat Ser Ile Asp	7709
999 gtg gt Gly Val Va 2165	t alig wax 1 Met liya	aat gaa c Aso Glu A 2190	jt tat caa ig Tyr Oln	tat gat azt Tyr Asp Asn 2195	aat caa cgc Aso Glo Arg	7757
the ago ca Leu Sar Gl 22 00	u làu gpu	tgt gag g Cys Glu G 2205	ի այր Թյո	tet eeg att Sur Pro Ils 2210	gat cat arg Asp His Thr 2215	7805
ggt ogt gt Gly Arg Va	a tta mat 1 Lau Asn 2220	ണു തുളേഷ Gln Gln I	t tac cat le Tyr His 2225	Tyr Asy Glu	tgg gga aat Tro Gly Aso 2230	7853

att meg ogg ott gat ast ack that oga gat ggt aag gaa acg gig gat The Tya Arg Lox Asp Asn The Tyr Arg Asp Gly Lys Clu The Val Asp 2235 2240 2245	7901
tat cat the agt cas goe gat cos act cas out att opt att see age Tyr llis Phe Ser Cln Als Asp Pro Thr Gln Leu lle Arg Ile Thr Ser 2250 2255 2260	7949
ger aan dag dag ata gag tta ag. Dat get get aat gge and dta add Amp Lys Olm Glu Ile Glu Leu Ser Tyr Asp Ala Asm Gly Am Leu Thr 2265 2270 2275	7997
ogt use gos ass ygg cas arg etc att tae gat can sat ast ege ttg Arg Asp Glu Lys Gly Glu Thr Leu Ile Tyr Asp Glu Asu Asu Arg Leu 2280 2285 2290 2295	804 5
gta can gto and get ogg tig ggc aat rig gig ige age iar reg tat. Val. Glo Val Lya Asp Arg Leu Gly Asu Leu Val Cys Ser Tyr Glo Tyr 2300 2305 2310	0093
got got the ear asa tha acc gra can git the god ast get acc git. Asp Ala Leu Asm Lys Lou Thr Ala Gim Val Leu Ala Asm Cily Thr Val 2315 2320 2325	8141
aat oga oag oot tet got toe gg! aas gtg ang aat ett om ttg ggt Asn Ang Gin His Tyr Ala Ser Gly Nys Vol Thr Asn Ile Gin Leu Gly 2330 2335 2340	8189
gat gas gag att act tog ttg age agt gat aag cea ogs att ggs ozt Asp Glu Ala Ile Thr Trp Leu Ser Ser Asp Lys Glu Arg Ile Gly His 2345 2350 2355	B237
can agt gor sag aat ggt caa tom gto tac tat caa tat ggt att gac Gin Ser Alæ Lys Asn Cly Cln Ser Val Tyr Tyr Gin Tyr Cly 11e Asp 2360 2365 2370 2375	82.R5
cat aac agt acg git atc goo agt cag aac gam aac gam itg atg got His Asn Ser Thr Val Ile Ala Ser Gln Asn Glu Asn Glu Ieu Met Ala 2380 2385 2390	6333
tta too tat aca cot tat ggo tit agg agt tta att too toa tta cog Leu Ser Tyr Thr Pro Tyr: Gly Phe Arg Ser Leu Tie Ser Ser Leu Pro 2395 2400 2405	8381
ggt tig aat gge gea cag git gat ees git ace gge tog tae tie tie Gly Leu Aen Gly Ale Gln Val Asp Pro Val Thr Gly Trp Tyr The Leu 2410 2415 2420	8429
gart and son this ords gith the and con gith other and age the cold age. Gly Amo Gly Tyr Arg Val Phe Amo Pro Val Leu Met Arg Phe Ris Ser	
2425 2430 2435	B477
	B477 B525
2425 2430 2435 for gat agt tog agt oot tit ggt ogg gga ggg att aac cet tet acc Pro Asy Ser Try Ser Pro Phe Gly Arg Gly Gly He Ash aro Tyr Thr	
2425 2430 2435 cor gat out tou out out out out out out out ou	ß525

Val Gly 11e Val Sor Leu Gly Ala Gly Ala Ala 11e Ser A1e Gly Leo 2490 2495 2500	
att get geg ggg gge get ttg ggg geg att get tet ace age geg ett Die Ala Ala Gly Gly Ale Leu Gly Ale Die Ale Ser Tur Ser Ale Leu 2505 2510 2515	B717
gra git act gog act gic att ggm ttg gct gcc gmt tog ata ggg att Ala Vai Thr Ala Thr Val Ile Gly Leu Ala Ala Asp Ser Ile Gly Iie 2520 2535 2530 2535	87 6 5
gog toa gon goa tha ing gaa aaa gat cog aaa ana tot ggg ata tta Ala Ser Ala Ala Leu Ser Glu Lys Asp Pro Lys Thr Ser Gly Ile Leu 2545 2550	8813
aat top att agt gog gga the ggg gtt tta agc tit ogt atc agc gcm Asm Top lie Ser Ala Gly Leu Gly Val Leu Ser Phe Gly lie Ser Ala 2555 2560 2565	8861
ath acc tit see tel ten etg gia ass ton ges one ant not tel eag The Thm Rhe Thm Sem Sem Leu Vol Lys Sem Ala Amp Sem Gly Sem Gln 2570 2575 2580	B909
ges gto age geg ggt gtt ato ggg tea gtg eet ett ges tit. ggt gaa Ala Val Set Ala Gly Val Ile Gly Ser Val Pro Leu Glu Phe Gly Glu 2585 2590 2595	8957
gtt get age ogt toe age age ogs tgg gat att gog tia tet tog aba Val Ala Sar Arg Ser Sar Arg Arg Tro Asp Ilo Ala Lou Ser Ser Ilo 2600 2605 2610 2615	9005
teg tig gge gen aat geg geg tet ete tet aog ggg ata geg geg geg Ser Leu Gly Ala Am Ala Ala Ser Leu Ser Thr Gly Ile Ala Ala Ala 2620 2625 2630	9053
geg git gra gad agi aai geg aai gea get aai ai: eig gga igg gid Ala Val Ala Asp Ser Aso Ala Aso Ala Aso Ile Leu Gly Trp Val 2635 2640 2645	9101
too tit ggt tit ggt goa gia tog ace and toe gga ata also gag oft. Ser The Gly Phe Gly Ala Val Ser Thr Thr Ser Gly The Ile Glu Leu 2650 2655 2660	9149
and topt and got tot gos our eat cat cag and top gas only eat too The Arg The Ala Tyr Ala Val Ash Gir Gir The Tro Gir Leu Ser Ser 2665 2670 2675	9197
toa goa gyt act tog gag gan gtg aag oot ata cyt tyt oto gtf; tra Ser Ala Gly Thr Ser Glu Glu Val Lys Pro Ile Arg Cys Leu Val Ser 2680 2685 2690 2695	9245
cae oge top ant ong ang rag tya atyttaacce teeteggyen gktysytton His Ary Trp Asn Gln (vs Gln 2700	9299
tomaacytti eymmatayta eeyyymmets titageemmi eyteemitym anoocytmat	9359
gtyttgegae gtogtttgae aatataaags ttetgogaam egaktegtta agtetooga	
aaaataacta ttaggogaca tttgogtogo otittttaaag gaactttato aggottacatt	
talaagaago tallitigitti togsoggsig tiggittoik: togsogsiaaaaa matagaggga	
salgstotcs aggytostas togitasity tamastatot patattatic postitatat	9599

greentgias tiectritai taittaatti taitgeatti getaegegaa alegeriiai 9659 aattitatti tiaatasatt attatteet esitaaseta aastaaatta tiictags 9717

<210> 2

<211> 417

<212> PKT

<213> Photorhabdus luminescens

<400× 2

Met Gln Any Ala Gln Any Val Val Ile Thr Gly Met Gly Ala Val Thr 1 5 10

Pro Ile Cly Clu Amp Val Clu Ser Cys Trp Gln Ser lle Ile Glu Lys 20 25

Gin Nie Arg Fhe Ria Arg Ile Glu Phe Pro Asp Ser Phe Iie Asm Ser 35 40 45

Amy Phe Pive Ser Phe Leu Ala Pro Asn Pro Ser Any Tyr Gln Leu Leu 50 55 60

Pro Lys Lys Leu Thr His Thr Leu Ser Asp Cys Gly Lys Ala Ala Leu 65 70 75 80

bys Ala Thr Tyr Gln Ala Phe Gly Val Asa Ile Sar 85 90 95

Pro Val Glu Tyr Tyr Asp bys Tyr Glu Cys Gly Val Ile Leu Gly Ser 100 105 110

Gly Top Cly Ala Ile Asp Asm Ala Cly Asp His Ala Cys Gln Tyr Lys 115 120 125

Gin Ala Lys Leu Ala His Pro Met Ser Asa Leu Ile Thr Met Pro Ser 130 140

Ser Wet Thr Ala Ala Cys Ser Ile Met Tyr Clly Lou Arg Gly Tyr Gln 145 150 155 160

Asm Thr Val Met Ala Ala Cys Ala Thr Gly Thr Met Ala Ile Gly Asp 165 170 175

Ala Phe Ghu Ile Ile Arg Ser Gly Arg Ala iws Cys Met Ile Ala Gly 180 185 190

Ala Ala Glu Ser Leu Thr Arg Glu Cyr, Asm Ilo Trp Ser Ilo Asp Val 195 200 205

Iou Asn Ala Leu Ser las Glu Gln Ala Asp Pro Asn teu Ala Cys Cys 210 220

Pro Phe Sur Leu Asp Ang Ser Gly Phe Val Leu Ala Glu Gly Ala Ala 225 235 240

Val Val Cys Leu Glu Asn Tyr Asp Ser Ala He Ala Arg Gly Ala Thr 245 250 255

The Leu Ala Glu Ile Lys Gly Tyr Ala Glu Tyr Ser Asp Ala Val Asp 260 265 270

Leu Thr Arg Pro Thr Glu Asp Ile Glu Pro Lys Ile Leu Ala Ile Thr 275 280 285 lys Ale The Ghu Ghu Ale Ghu Ile Ser Pro Lys Asp Ilo Asp Tyr 1le 290 295 300

Aan Ala Mis Gly Thr Ser Thr Pro Lou Asa Asp Leu Tyr Glu Thr Oln 305 310 320

Ala Ile lys Ala Ala Leu Gly Gln Tyr Ala Tyr Gln Val Pro Ile Ser 325 330 335

Ser Thr Lys Ser Tyr Thr Gly His Leu Ile Ala Ala Ala Gly Sor Phe 340 350

Glu Thr lle Val Cys Val Lys Ala Leu Ala Glu Asn Cys Leu Pro Ala 355 360 365

Thr Leu Asn Leu His Arg Ala Asp Pro Asp Cys Asp Leu Asn Tyr Leu 370 380

Pro Aen Glu Idis Cys Tyr Thr Ala Glu Pro Glu Val Thr Leu Aen Ile 385 390 395 400

Ser Ala Gly Phe Gly Gly Ris Ash Ala Ala Leu Val Ile Ala Lys Val 405 410 415

AUG

<210> 3

<211> 253

<212> 180

<213> Photorhabdus lundnescens

<4**00**> 3

Met Glu Asp Ile Glu His Tro Ser Asn Phe Sor Gly Amp Phe Asn Pro 1 5 10 15

ito Val Gin Gly Met Leu Ser Leu Leu Tyr Val Arg Gin Gin Phe Ser 35 40 45

Gin Lev The Ser Ala Pas Thr Thr Gly Ile Lev Asa Ile Asp Ala Ser 50 55 50

Phe Arg Glo Tyr Val Tyr Thr Ala Leo Pro Kis Glo leo Arg Ile Aso 65 70 75 80

The Lys Aso Lys The Phe Lys Leo Glo Aso Pro Sar Lys Glo Aso The 85 90

Leu Phe Cly Asm Thr Ser Val Glu Asm Thr Met Glu Ser Ile Glu Asp 100 105 110

Trp Ile Val Gln Asp Aen Cys Gln Lys Leu Thr Ile Thr Cly Glu Glu 115 120 125

Vai Cys Glu Lys Tyr Ala Val Pho Arg Tyr Tyr Phe Pro Sex Vai Thm 130 135 140

Ser Ile Cly Tro Phe Leu Asp Ala Leu Ala Phe His Leu Ile Ile Asm 145 150 155 160

Ser Thr Gly Phe Leu Asn Phe Glu His Tyr His Phe Asn Gln Leu Gln 185 170 175 And Tyr Leu Ser Glin Ser Phe Thr Leu Kis Thr Gly Glin Ala Ile Lys 180~ 180~ 185~

Ile Arq Lys Glu Ile Vol Asn Scr Thr Vol. Leu Leu Ser Ser Pro Asp 195 200 205

The Cys Wal Glu Leu Asn Pro Pro Leu Leu Ile Lye Asn Gly Asp Lys 210 220

Amp Tyr Ile Arg lie Whe Tyr Tyr Arg Cym Leu Tyr Amy Lym Pro 225 230 230

lle Hoe Val Ser Lys; Thr Ser Ile Ile Ser Lys; Met Lys 245 250

<210> 4

<211> 186

<212> FRT

<213> Photorhabdus luminescens

<400> 4

Wet Asn Val Leo Glo Glo Gly Lys Val Ala Ala Leo Tyr Ser Ala Tyr 1 5 10 15

Ser Olo Thr Glu Gly Ser Ser Trp Val Gly Asn Leo Cys Cys Whe Ser 20 25 30

Ser Asp Arm Glu His Leu Pro Ile Ile Val Asp Gly Arg Arg Phe Leu 35 40 45

The Glu Phe Val The Pro Asp His Leu Leu Asp Lys The Val Lys Pro 50 55

Arg Val Phe Asp Leu Asp Ile Ash Lys Glo Pho Lau Lau Arg Arg Asp 65 70 75 80

His Arg Clu Ile Acm Ile Tyr Leu Leu Cly Clu Cly Asm Phe Met Asp 85 90 95

Arg Thr Thr Pap Lym Asn Leu Phe Glu Leu Asn Glu Asp Gly Ser 100 105

Leu Phe Ile Lys Thr Leu Ary His Ala Leu Gly Lys Tyr Val Als Ile 115 120 125

Asn Pro Ser Thr Thr Glo Whe Tie Dhe Whe Ala Glo Gly byz Tyr Ser 130 140

Glu Phe Ile Met Asn Ala Leu Lys Thr Val Glu Asp Glu Lou Ser Lys 145 150 155 160

Nrg Tyr Arg Val Arg Ile Ile Pro Glu Leu Gln Gly Pro Tyr Tyr Gly 165 170 175

Phe Glu Lou Asp Ile Lan Ser Ile Thr Ale 180 185

<210> 5

<211> 258

<212> PRT

<213> Photorhabdue luminescens

<400> 5 Met Olu Lys Lys Ile Thr Thr Phe Thr Ile Glu Lys Thm Asp Asp Asn Phe Tyr Ala Amn Gly Arn His Gln Cys Met Val Lys Ile Ser 15 20 Val Leu Lya Glu Glu Tyr Arg Axu Gly Axu Tro Ho Lya Leu Ala Leu 35 40 Ser Glu Ala Glu Lys Arg Ser The Cln Vol Ala Ala Len Ser Amp Ser 50 55 Leo lie Tyr Aep Gin Leo Lys Met Pro Ser Gly Trp Thr Thr Thr Asp Ala Ang Asn Lys Phe Asp Len Gly Len Len Asn Gly Val Tyr His Ala Asp Ala Phe Ile Asp Glu Gln Val Thr Asp Arg Ala Gly Asp Cym Cym 95 100 105 105 Thr Asn Ghu Aen Tyr Gln Aen Ser Val Lys Ser Val Pro Clu Yle Yle Tyr Arg Tyr Val Ser Ser Asn Arg Thr Ser Thr Glu Tyr Leu Met Ala Lys Met Thir Pixt Glu Asp Thir Asp Gly Lys Arg Thir Leu Thir Thir Ash 145 150 155 Met Ser Val Gly Asp Giu Val Phe Asp Ser Lye Val Leu Leu Lys Ala lle Ale Pro Tyr Ale Ile Asn Thr Asn Gln Leu Wis Clu Asn Ile Asn Thr Lew Pho Asp Lys Thr Clu Olo Pau Tha Lys Sor Asp Thr Hie Hie 195 200 205 Gln lle lle Asn Leu Tyr Arg Tro Thr Leu Pro Tyr His Leu Arg Ile Low Olu Gly Asn Asp Ser Thr Vel Asn Arg Ile Tyr Vel Leu Gly Lyz 225 230 235 Glu Pro Ser Asn Asp Arg Phe Leu Thr Arg Gly Arg Vel Phe Lys Arg 240 Gly Thr His Met 255 <27,0> 6 <211> 1584 <212> PRT <213> Photorhabdue luminescens <400> 5

Met las Ala Thr Asp Ilo Tyr Ser Asa Ala Pho Asa Phe Gly Ser Tyr 1 5 10

Ile Asn Thr Gly Val Asp Pro Arg Thr Gly Gln Tyr Sex Ala Asn Ile 20 25 30

- Amn lie The Leu Arm Pro Amn Asm Val Gly Asm Ser Glu Cln Thr 35 40 45
- Leu Ser Leu Ser Phe Ser Pro Leu Thu Thur Leu Ash Ash Cly Phe Gly 50 55 60
- Ile Gly Trp Ang Phe Series) Thr The Leu Asp lie Lys Thr leu Thr 65 70 75 80
- Phe Ser Arg Ala Asm Gly Glu Gln Phe Lys Cys Llys Pro Leu Pro Pro 85 90 95
- han han han hap ben Ser Phe Lys hap Lys Lys ben Lys hap ben hmg 100 105 110
- Val Tyr Lys Lou Asp Sox Asm Thir Phe Tyr Val Tyr Asm Lys Asm Gly 115 120 125
- Ile Ile Glu Ile Leu Lys Arg Ile Gly Ser Ser Asp Ile Ala Lys Thy 130 135 140
- Val Ala Lou Glu Pho Pro Asp Gly Glu Ala Phe Asp Leu Ilo Tyr Asn 145 150 155 160
- Ser Arg she Ala Leu Ser Glu Ile Lye Tyr Arg Val Thr Gly Lye Thr 165 170 U75
- Tyr Leu Lys Leu Asu Tyr Ser Gly Asu Asu Cys Thr Ser Val Glu Tyr 180 185 190
- Pro Asp Asp Asp Asp Ile Sur Ala Lys Ile Ala Phe Asp Tym Ary Asn 195 200 205
- Amp Tyr Leu Ille Thr Val Thr Val Pro Tyr Amp Ala Ser Gly Pro Ille 210 215 220
- Amp Ser Ala Arg Phe Lym Met Thr Tyr Glo Thr Leu Lym Gly Val Phe 225 230 235 240
- Pro Val Ile Ser Thr Phe Arg Thr Pro Thr Cly Tyr Val Glu Leu Val 245 250 255
- Ser Tyr Lya Glu Ash Gly His Lys Val Thr Asp Thr Glu Tyr He Pro 260 265 270
- Tyr Ala Ala Leu Tir Ile Gin Pro Gly Asn Gly Gin Pro Ala Val 275 280 285
- Ser Lys Ser Tyr Glu Tyr Ser Ser Val His Aso Phe Leu Gly Tyr Ser 290 295 300
- Ser Gly Arg Thr Ser Phe Amp Ser Ser Gln Amp Am lea Tyr Lea Val 305 310 315 320
- Thr Gly Mas Tyr Thr Tyr Ser Ser Ile Glu Arg Val. Leu Asp Gly Gln 325 330 335
- Ser Val Val Ser Val Ile Glu Arg Val Phe Ash Lys Phe His Lau Met 340 345 350
- Thur Lys Glu Ale Lys Thur Glu Asp Asp Lys Arm lie Thur Thur Glu Ile 355 360 365
- The Tyr Asn Glu Asp Lou See Lys Ser Phe See Glu Gln Pro Glo Asn 370 275 380

1/xx 3 8 9	Gln ,	CЫ	Pro	Serv	Л гд 390	Val	Leu	Tiu	Arg	Tyr 395	Tîne	A61)	Ilc	Gln	Tha: 400
A≊T.	1tm	Ser	Arg	Glu 405	Glu	Tear	Val	Asn	Ilc 410	Lys	Ser	Авр	Asp	Ттр 415	_
Ąşr	Mi	. इस्ता	Leu 420	Ile	Tìư	Glu	Thu	Ser 425	Gĵλ	Ile	Gla	Lys	Glu 4 30	Tyr	Val
Ίγτ	TYX	Pro 435	Val	ÀSO	Gly	Glu	Gly 440	Asn	Ser	Сув	Pro	Ala 445		Pro	Leu
Gly	Phe 450	Ser	Arg	Ehe	Leu	lys 455	Ser	Val	Thr	Glu	Lys 460	Gly	Ser	Pxo	As ₁ >
Ala 465	Ala	Gln	Ser	Val	Ala 470	Aen	lу́е	V≥l	Ile	H1s 475	Тут	The	Тут	G i n	Lye 480
Phe	Pro	Tìur	Phe	ቸስታ 4B5	ĠĴŸ	Alæ	Туr	Vail.	I.թթ 490	Olu	lyr	Val	Ser	Lye 495	Va).
Sar	Glu	The	11e 500	Asp	Asn	ГАЗ	IJe	Ala 505	gTA	utur	Fire	Sen	Тух 510	Val	Asn
		212					520					525			
Met	Agn 530	A523	Gl'n	(Alm	Than	Va1 535	The	The	Pho	lys	Tyr 540	CJu	ТУТ	Ser	Glu
5er 545	Glu	Met	Thr	'Itar	Asto 550	Ala	Thar	val	TŶŒ	Gly 555	Phe	ÀSQ	Gly	Ala	Нів 560
Net	Glu	Ser	nys	Aen 565	Va.J.	Thu;	Ser	J.le	Тут 570	Thr	Kie	Arg	СĴЪ	Leu 575	Лту
М	Val	yab	Va1 580	ঠিকা	Ris	Val	Πe	Thr 585	Авр	Gln	Ser	Ty1:	λ ερ 590	Legy	Leu
હાઝ	Ang	I) e 595	Tru	Cly	Gln	Ile	11e	Asp	Pro	СĵУ	Tìu	Ala 605	Ang	Glu	De
Lуe	Arg 610	Aen	TYI	Val	Tyx	Gln 615	ፒሃድ	Prv	Gly	Gly	Авр 620	Glu	Aso	λε γ)	Phie
Trp 625	Pro	Val	Mest	IJe	Glu 630	Val	Аер	5er	Gln	Gly 635	Val	Arg	Arg	Lys	11m 640
His	Tyr	Asp	Gly	Met 645	cly	Arg	Ile	೧/೯	Ser 650	Πé	Clu	Glu	Gln	Asp 655	cps.
увъ	СĴУ	Alz	Этр 660	Gly	The	Sor	СЛУ	Ile 665	Тух	Gln	Gly	אלני	тут 670	Ang	Lys
Val	Leu	Ala 675	Arg	Gln	Tyr	ÀЭ	Val 680	Leu	Gly	Gln	Leu	Ser 685	Dye	Glu	Ile
Ser	Asn 690	ÀSO	Тгр	Leu	dτ	Aen 695	Leu	Ser	Ala	Assn	Pro 700	Lou	Val	Arg	L€n
Ala 705	Thr	Pro	Leu	Val	รกะ 720	Thr	Lyz	MY	Tyr	185 715	'Iyr	ĄSĄ	Gly	Ттр	Gly 720
Astr	Leu	ፕ ሃ ኦ	Sear	Thr	Glu	ባን포	Ser	ሊ ሞን	OLY	Arg	Ile	ශ්ය	Leu	Gl ប្	IJŧŧ

				725					730					735	
HiB	PSP	PTO	11e 740	Tìm	Arq	Thr	Ile	Thr 745	Gln	Сlу	Val	Lys	Gly 750	æ	G7y
Met.	Lou	Asn 755	ī.le	Gln	Glin	Asn	Asn 760)	Phe	Glu	GJIJ	Pro	Ala 765	Seor	lle	Lye
Ala	Val 770	zy r	Pro	Asp	Gly	Thir 775	Πle	Tyr	Ser	TÎM:	Arg 780	Time	Тут	Arņ	ΊУΤ
Asp 785	Gly	Etvo	Gly	Агд	Thar 790	Val	The	Glu	Thr	леқі 795	Ala.	Glu	Gly	}tie	800 УЈЧ
Thu	Gľu	Ile	GLy	Tyr 805	Мър	Val	Phe	Asp	Лгу 810	IJρ	۷al	Lye	Dye	Thr 815	Lou
Pro	Asp	ΟĴУ	Thr 620	Ile	Len	Glu	Ser	Ala 825	Tyr	aLA	Ser	Phe	Ser 630	His	Glu
Glu	Lena	11c 835	Ser	Ala	Leu	Asm	Val 840	Asn	Gly	Ή τ	Cla	Leu 845	Gly	УĴя	Lou
Val	Ty1 650	Asp	Gly	Leu	GЉ	Arg B55	Val	Lle	Sear	Asp	Thr 860	Val	Gly	Дy	УЛĒ
Lys 665	Tìm	ദിധ	Tyr	Leu	Tyr 870	Ģly	Pxo	Gln	Gly	Азр 875	Dya	Pro	Re	Glu	Ser 880
Ile	Thr	Pro	Ser	и16 805		Lys	Gln	Aen	Me t 890	Asp	Тут	Zen	Тут	Tyr: 89 5	
Ġĵу	Ser	Val	Met 900		Lys	Ehe	Thr	Thr 905		Thr	<i>P.</i> 9p	Gln	Gl n 910		Phe
ĀĽĢ	Тут	His 915		Lys	Thr	gly	ਾਨਿ: 920		Leu	. Ser	Ala	Ser 925		Gly	Val
Ser	Gln 930		P .20	Тук	Ser.	Тут 935		Pro	S€T	Gly	Va) 940		G)r	Arg	(GIV
9en 945		. Leu	Arg	ASE:	950		: Pro	. IJ∈	: Ser	Ser 955		Gle	. ጋ እፕ	Len	1yr 960
'Ita	19et	. Sen	Gly	965		Gln	Arg	His	: Lye 970	. Asp)	Ser	Phe	Gly	(119 97 <u>5</u>	
)ប <u>់</u> ន	: Val	Tyr	980		ASE	A)a	Gùn	Gly 985		Leu	Val	. Lya	ि प्राप्त 990		G)r
Asp	Ala	Gln 999		Ala	r it a	Ehe	: Glu 1000		Ası) Asın	. Val	. Gly 1005		leo	lle
Tha	Thi 1010		Thi	T.NS) Asp	Thi 1015		Sei	Len	ı Ser	1020 1020		VS)	Tin	Lys
11e 025		l Tyn	· Act	Ala	1030		Arg	(Glu	ı Ile	≥ Lys 1035) Seo	C Len,	Ile	9 S€1 184(
Ast	e lêtne	9 S eo	116	Gle 1045		1 3 1 c	: Tha	: Le	. <i>9</i> உ 1050	r Tyn	Thi	i Lyz	aen	1055	
Ile	: Sei	c Glr	1. A rg		rtTr ≘	Ser	. FJe	АБД 1069		y Val	. Val	. Y ⁄≘1	Lyt 1070		3 G1:

- Arg Tyr Gln Tyr Asp Asn Asn Gln Arg Leo Ser Gln Tyr Gln Cys Glo 1075 1080 1085
- Gly Giu Gln Ser Pro Ile Asp His Thr Gly Arg Val Leu Asn Gln Gln 1090 1095 1100
- Ile Tyr His Tyr Asp Gln Trp Gly Asm Ile Lys Arg Leu Asp Asm Thr 105 1110 1115 1120
- Tyr Arg Asp Gly Lys Glu Tur Val Asp Tyr Ris Phe Ser Gln Ala Asp 1125 1130 1135
- Pro Thr Glm Leu Ile Arg Ile Thr Ser Asp Lys Glm Glm Ile Glm Les, 1140 1145 1150
- Set Tyr Asp Ale Asn Oly Asn Lea Thr Ang Asp Glu Lys Gly Gln Thr 1155 1160 1165
- Len Ile Tyr Asp Gln Asn Asn Arg Len Val Gln Val Lys Asp Arg Len 1170 1175 1180
- Cly Asn Leo Val Cys Ser Tyr Gln Tyr Asp Ala Leo Asn Lys Leo Thr 185 1190 1195 1200
- Ala Gln Val Leu Ala Asn Gly Thr Val Asn Arg Gln His Tyr Ala Ser 1205 1210 1.215
- Gly Lys Val Thr Asm Ile Glm Lev Gly Asp Glu Ala Ile Thr Trp Lev 1220 1225 1230
- Ser Ser Amp bys Glin And life Gly Ris Glin Ser Als Lys Amn Gly Glin 1235 1240 1245
- Sor Val Tyr Tyr Gln Tyr Gly Ile Asp Ris Asn Ser Thr Val Ile Ala 1250 1255 1260
- Ser Glin Asm Gliu Asm Gliu Leu Met. Alsa Leu Ser Tyr Thr Pro Tyr Gly 265 1270 1275 1280
- The Arm Ser Leu Ile Ser Ser Law Pro Gly Leu Asn Gly Ala Gln Val 1285 1290 1295
- Asp Pro Val Thr Gly Trp Tyr Fhe Leu Gly Asm Gly Tyr Arg Val Phe 1300~1300~1305~1310
- Asn Pro Val Leu Met Arg Ehe His Ser Pro Asp Ser Trp Ser Pro Phe 1315 1320 1325
- Gly Arg Gly Gly Ile Aso Pro Tyr Thr Tyr Cys Gln Gly Amp Pro Ile 1330 1335 1340
- Asn Arg The Amp Leu Am GDy His Leu Sor Ala GDy GDy The Leu GDy 345 1350 1355 1360
- Ile Val Leu Gly Ala Ile Gly Ile Val Gly Ile Val Ser Leu Gly 1365 1370 1375
- Ala Gly Ala Ala Ile Ser Ala Gly Leu Ile Ala Ala Gly Gly Ala Leu 1380 1385 1390
- Gly Ala Ile Ala Ser Thr Ser Ala Leu Ala Val Thr Ala Thr Val Ile 1395 1400 1405
- Gly Leu Ala Ala Asp Ser Ile Gly Ile Ala Ser Ala Ala Leu Ser Glu 1410 1415 1420

```
lys Asp Pro Lys Thr Ser Gly Lle Leu Asp Tro He Sor Ala Gly Leu
Gly Val Leu Ser Phe Gly Ile Ser Ala Ile Thr Phe Thr Ser Sor Leu
               1445
                                   1450
Val Lys Ser Ala Arg Ser Gly Ser Gln Ala Val Ser Ala Gly Val Ile
City Ser Val Pro Leo Glo Phe Gly Glo Val Ala Ser Arg Ser Ker Arg
Arg Tro Asp The Ala Leu Ser Ser The Ser Leu Gly Ala Asa Ala Ala
                       1495
                                           1500
Ser Len Ser Thr Gly Ile Ala Ala Ala Ala Val Ala Amp Ser Asn Ala
                   151D
Aso Ala Ala Ast Ile Lou Gly Top Val Ser Five Gly Five Gly Ala Val
                                  1530
Ser Thr Thr Ser Cly lle lle Glu Leu Thr Arg Thr Ale Tyr Ala val
           1540
                               1545
                                                   1550
Aso His Glo The Top Glu Leu Ser Ser Ser Ala Gly Yor Ser Glu Glu
                          1560
Val iya Pro Ile Ary Cya Lou Wal Ser His Ary Tro Asn Gin iya Gin
                       1575
<210> 7
<211> 18
<212> DNA
<213> Artificial Seguence
<220>
<223> Description of Artificial Sequence:oligonacleotide
<400× 7
acacagcagg ttcgtcag
                                                                   18
<210> 8
<211> 18
<212> TBAA
<213> Artificial Segmence
<223> Description of Artificial Sequence:oligonuclootide
<400> 8
gycegeagus ctceachd
                                                                   18
<210> 9
<213> 20
<212> DNA
<213> Artificial Sequence
د2200>
<223> Description of Artificial Sequence:oligonurlectide
<400> S
```

```
attgatagra ogeggegene
                                                                   20
<210 - 10
<211.5 22
<212> DNA
<213> Artificial Sequence
<220>
<223> Description of Artificial Sequence:oligonucleotide
<400> 10
tiginacqiq qaqeeqaset qg
                                                                   22
<210> 11
<211> 37948
<212> 1MA
<213> Photorhabdus luminescene
<220>
<221> CDS
<222> {35171)..(18035}
<223> orf5
<220×
<221> CDS
<222> (23768)..(31336)
<2235 hph2
<220>
<221> CDS
<222> (31393).. (35838)
<223> onf2
<400> 11
igitegotega orgilggagat tatgeotato giongitaga acgagacang otomatgang 60
ccaagatgtg gtatatgcas graetgcate tgllaggoga tamacetest ctategthea 120
gitcagagig gagraaarog agittaggog gogcigxxyg aarggaaaga raaaagcaar 190
acgrocasgo aatggooget, ctgcgscaag gtgatgttag teggcaraac aacccgscag 240
atottitott goomnaggid malgmagigs igomaamota tiggmaamaa iiggmaddan 300
ggstgtataa ootgogtest escetoarta tigsengenea acopotacat etgestalet 360
acgotacaco ggoagatoca asagosttao ttagogoogo tytogotago teogaagyty 420
ggytagetet etcacageca titatgtoae tgtggegtit emacaeaby etggaaaacy 080
egogbygtat gytcagtoag otosetozat tegyptotec yotaczazat attatogase 540
gramagnatur gymmyettta memogetet tgragmatem agengemen etgatattym 600
ctesteteng entacempase assacenter caraceters; tgegyaxaxx atempaetes; 660
azasatowa sycogypycy czatenogot ttymczycta czasaegtta (wcymcymae 720
statomaly: gyglgmaam ogggetatag cuttgomige eteegitget ggeekomgem 700
ntgecetycz agostożegt etgycygycy etgogethya tetgycycoc aaratititog 840
ytalagetga iggoggiage egitiggigag egatigaxya agegacagyi aaigitaigg 900
```

estimicogo cagigigatg aacamogasg oggataasai cagmmaguda gaayootaho 960 ggggtogeeg teaggaatgg gaaatocage gtaatestge ogsagesgag ataasacaga 1.020 togatectea acticastea etggeagtse geogtgaage egegystatte eagamageea 1080 gertsasser cosersges cagacidate richatiges bilosigna ogiasetica 1140 gtastcaago gttgtactac tggctacgog gtcggckagu tgctatttac ttccaatttt 1200 acceptings of agency to be a second of the contract of the con agacogogo nagotitato naaccoggog ectggcaggg aaccomatgog ggibbantigg 1320 ctogtesaso ct/gateoby autotegogo anategonaga teconstitis agetegogate 1380 ascgryptick ggaegtogaa opgacoatti cattymugoz acartaigga gcactgrozg 1440 assautogut tastitayoo acaaggatti etacootgot agraggiggi acaanigaci 1.500 creatigatga teatroxytt acattagana acquecanot tagigecana sicietetgi 1560 caggiratgia attagataat gartachag alagonacago ogtoggmaac attogacyca 1620 trassparat captgreace troccapeer tottaggent statezogat graceageta 1680 ttotgtocta oggaggaagt gaaateggat tagetgaaag otgtaaatea otggogatet 1740 ctoatgrout caatgaragt getcaatter agttygatti (aaccatggt aagtteetge 1800 cytthywwgg guttgogatt ascgatecty goscattysc autozyttto coccaatgog 1860 actytoware amemamacat gitgeogeri tigegtysis tiatictyca tattogetat 1920 scratcogem asiamocaco towattamat necammana gyetecima egggyootga 1980 actiticacy sataiaiant actoscapto igotototti accigiciga ogcingital 2040 aacagagata titectitic icqtqqqqq celquostac tataaaatai caacccicti 2100 ettittrata atatgeaata lytaacaat yeaattatit eatitagita tigitaacta 2160 gitatatiae tiatquiqte attataeett tigttetige atcacaatag cestiteset 2220 aaataataac gttgtgaaat agtbgatagt tzaatggtgt tittettieg cogitatlit 2280 caacocaatt teagacegek abbagaegit acctgtgttg cetttgtttt gategatata 2340 estaapotta titatatoos ogytaniosg arregostas atgititati tarciaacat 2400 ttaaanggaa tasacatgaa cacactcaaa trogastatg anaacgcytt agtagcagyt 2460 titaataate taacegatat Ulgiombete teithigmog aachtegeas aaaagigaag 2520 gacasactot calggtosca gaccossago tistatotitg asgoscagos agigoscana 2580 gaceal/Clon tocatgaage ocatationig aaacgegoda 60cctoatti acazagigog 2640 gurostotte cootenaac acctuatect gaccaguese ettakaatag capatiteese 2700 astrogorus guasatatzo agroccaggo guasattiett edatgiitu teotgoggot 2760 tatttagotg aactttatog toxogoxogg aatttacatg stgassattc tatttatcat 2820

ttogatacac googtoogga totaaastos ttogtootos goosgaassa tatggatacg 28	80
pagattteen eactttetet gtetaatgae ätyttgetag appgraftag mactetgite 29	40
asgyacasgo tgotgguygo totgasgast attaastoto tgtxcasgas ogagotgotg 300	00
geggettiga agaztattaa acetetgice aaggaegate fyelgowgge tiligaagaat 300	60
attamaeete igiocamaga ogaicigetg gaggettigs mysatatiam meeteigiee 312	20
aaggaagate tgetggagge titgmagmat attamaeete tgtegmagga egatetgetg 310	80
gaggethtga agsabablee acutetytet aaggaegate tgetggaggn httgaagaat 324	40
autaaacoto Egicomagga ogaioligolig gaggoutiga mgaatattaa accidigioo 330	00
aaggacgate tacaggaatg tattgaaatt etattcaate tagacageca caetaaagta 330	60
atgassgryt tatenasitt orgegittet gyratgatge catalescya byrttatgaa 341	20
agogtgogta aggttgttca attacaggot coggletttg aacacgttgl tagtacatca 340	ВО
ctagaaacga ctatogatga actamamtat campettett tgttggammi tamttettet 35	40
giologosta ambiatitas tatetigast gaagaaalta stasaatsaa igsaagaagi 360	00
etetatgagg aanattitigg taatattaan eettetetaa taggamanee ggaatatetg 366	60
esaagttatt sceek/tong tgatgaagag titagagatt tmaktamast magaamtata 37	20
ettettecag aagsagssal agsasttaet patettgeat egeglactee tagtacassa 378	ВО
cagactates aaaateetga kkalogiget elattuaaaa tiaataagti tattogieta 386	40
ticaaageta taaacttate acegacygta kiaagiggaa teeteegeag catcageaca 390	30
gazttokata tomatamaga matattacan masatolitto gigitamata diministrona 390	50
ogttätegta tigacautga gastgeatta atastatges egytassaat tisassatai. 40%	20
alceaugeon geosuctoes tougittigat ogittatitis autocoocaa actgasiggi. 400	80
coagattitt cogtomatgg tactomgant attgatting coctangong taccascmer 414	40
tyyaataaaa castacttaa acyteettit sacetegate atatetesti saatopaeta 420	OC.
ctamaaatta comatoogyt caatactacc gaaatgataa ctaatgatat agagaatett 420	60
ixitoatutol ataggaczas attactggca gatatorak: agttazotat tgatgaartg 432	3 0
gggttactgt tggsagguat aggtanagga acaaccaatt talutgagat tactoctgac 438	SI)
amicogytta etetamitea emmantetai geigieseta geiggeiaeg imesemmas 440	10
tygagtgiet aleagitytt fælgelged, echydiaest staaceamed Coleecoog 450	30
gaaatacgga atttactgga taccgtotac aatggottgc aagattijoaa taagaagatg 450	30
ttgaaagetg aagaagateL нужужжжих: мяжжидлаат tgcagegygc сихоролилик 462	30
стуравават тоссуудава ссвуствень ситовующ всердавава вустинующи 460	3D
agaetgaata aagetgaaga gaeecaegaa aaageegaga aaaacetaga tgaggteagg 474	113
assautotgo cazzagoost steteettat stegeng \exp etetgesatt seestetges 480)(1

cabgrogodsi attocstant cerchopyca gataalxiigg asucceptoat aggasasatg 4860 acageggasa aattatggaa etggtigegg aaaasteeeg tlauggetes acetgaatte 4920 casaaacaag ctgaacctgt ggtocagtat tgocageggg bggcacoact agegttgatt 4980 taccytteta cogyectias cyassaesee tiasyteigi litybynoona gecycasese 5040 tttgttatta aaaocaaago aoccgaaaca actgaaacaa caccegoaca tgaogtatca 5100 acactaatyt cactaacgeg tittactgac tyggitaact cactagguya aanogectet 5160 kototactaa oogaatttaa aaaagoaaca ttaacagcag aactatigge (aaagctatg 5220 aatritigata aaaaleteet gaagemagee magaitemagg caasametigs tiitoteemme 5280 typocatota togaczacot attyczytyg attaacstot cacytoaatt gazcatotot 5340 concengges bittoczonet gastonagta tigacognag amesicones tametataes 5400 coateggass amprogetge gatattazee geogggetgg acamecasas gaetzaegee 5060 ctacatgcgt ttctggetga gkctcgcagt gctgcgttaa gcacatacta (atttattct 5520 cataancaas aagatogaga ayosegaaas (atacagtaa ttaaaaacog tgsbgutota 5580 tatosatace tattgatega taaccasgtt tempcogeca ttaaasetac agagateget 5640 gaagetateg etagtateca actgiatati sacryogogi tgaasaatat ggagggagat 5700 actytomesa ytytomeesay eegetemite tieseematt yyymtäämiä essimmäese 5760 tacagosott gggcoggtat ggctaaacte cittactale cagagaatta csicgaiceg 5820 acyctacy(a bbygycagae znasatgatg gatacyttyc tycaatocat cagecaaage 5880 castteasta togataccol; agoagatoco titazatett acctaensto attogaacao 5940 gtggstaate tggaaateet cagegodkan oatgasaasa ttaataatga texaggatta 6000 acttacttta teggaegtag taasseegas gügaateaat attattggeg cagtgtggst 8060 caraataaat ocapogaagg taaalikooco gotaatgoot ggagtgagty gtacaasatt 6120 gattgtecaa ttaaccecta casagatact attcaccogg taatttteca atctogoctg 6180 tatottatot ggotggaaca aaamaayyoy sotamakongg amggtgataa sacogootog 6240 ggttallatt algesotgsa attagegsat atcegttatg ecagomestg gaataceexa 6300 gtraccutte augumanca massatatro gattiasair teggammatas macacotega 6360 ctttactgct caagetttea aggeagagat gaattgetgg tgatgtttta taaaaaacaa 6420 qatçasttas alcastacae associata eesstassag geolatatat caettecast 6480 etglolitoka egymantgan acetgaaaat caraaacets ergettataa acagiitigat 6540. antagingus kinkhtgetet czatantogo tatgoagaaa gotaceasat cocttoatoa 6600 gtazatagtz atazogytta Egattyggyk yatgyotato tgagtatgyk ytktygcygk 6660 aatatttoag coatoaaact ggagtootos toagstaagt taaaactoto soosaggilla 6720

aganttatte efeatggact tgtaggeoga caacgossee aatgcaacct gatgaagaaa 6780
tacogorage tiggigataa attiatiati talaciacic taggiatiaa ecceaataat 6840
tigtogaata eABBBITCat ttaccotgtt tatcagtata gigggaacac taccaataat 6900
родовардос gt.ctgctgtt ttategaqaa əgtactacta actttgtaag əgcctggttc 6960
octamoetti; cetetggete temmanatu teememme etggeggtga emitmegtggt 7820
$\omega *: \text{Caligate}$ atatiqataa casacatagi gaegaigite catitaaaca atatitciai 7080
atggatgace acggtggtat tgacaetgal gthloaggga tattatetat taatacgaac 7140
attaatoatt comeagtinu agtaatagtg amagcogaag gtatoacaga gcommotitit 7200
gtagogagog амансыglas lighteemiser satrogam; gebbegsaga aatgaattat 7260
cogtitacog agentigaaat agstatetee acactgaesi tikkataataa tgaagcaagt 7320
allogatales cottistogo attigotgag aastiigsog staatsguas tgatogiaso 7380
Upggogeag ascatticeg tattegtati atcasasasg eggassetgs Unstgeootg 7440
acectgrace atsatgrasa egggegeaa tatatgragt ggggaaacte ttgtattog: 7500
ettastadge tatttgeceg benættnætt egeogrageen acgeggggat agstactatt 7560
tigageatgg acaetcagna taiteaggaa cetaaattag gagaagatte teetgatget 7620
atgyaaccaa tgyacttoma oggogocaac agostotatt tetgygaact gttetactac 7680
accessatge tgattgeten acgststgetg cacquacema actsogstga ggetaecegt 7740
topotessat abgudiggsa uccatooget tatattetoa abgetoaaat qoaacattac 7800
ogetggaatg tengeceste scangaagse antagtligga acgstgatoc gttggatten 7860
titgateetg ataccatage teaacatgst coaatgcart acasagtuge coccittatg 7920
egeacectag atetyttgat egaacgggga gattacgect ategocaabb ggagcgagac 7980
acactegety aagecasaat ytyytataty caggcactym atmiattagy tyataaacet 8040
ratetaccae teagiticage aiggealight compagniag wayaggooge agetetigam 8100
assessesyy caratgross agasebayra gathlacgar auggactico taraterara 8160
gggtctaaag atgmaatcaa aacagatett tteetgeege aagiraxxga agtgotgetg 8220
agetactgge agamactaga acasegging istascence godstascent chotathgat 6280
ggtcasectt tacatttgcc tattttegca anaceagoag alumaaaago gotgetoago 8340
groupshipting computations aggregated amborbocat capacitatal atcaginging 8400
ogithteceka etalagobaga azaogeoogi agizingisa gisageiaae eesatiogga 6460
trescatter amountattat ognoceptena gotgoggagg esttaaacse getetteesa 8520
anteage/99 ca90aclgat attgaccaat etrogeatae apparaasse estensagag 8580
ctagatacty ಸಿತಿಕಿಕಿಸಿಕಿಗಳು actagasaan ಕಾರ್ಯಭರ್ವದ gaacccagte gegttitgat 8640
agetacagua mattotacga tgaagacato eaogegegog aazaacagge aatggegtig 8700

cgtgetteeg	tugctggcat	ctctacagoo	cticaagcat.	cacatotyge	99gogragca	8760
cttqatetgg	ogoCCGaaceat	ctteggct te	getgatggtg	geageegttg	gggggagstr:	H820
geocaageoa	caggtautgi;	m igga gitc	teegeeagtg	ttatgaacec	ന്യുഴുത്തുട്ടേ (.	8860
assatosgoo	aatotgaago	r:f.eccytcgg	cgtcgtcagg	aatgggaast	ticagogitaat:	8940
aacgoogagg	ದಾರಾವರು.ಗೌ <i>ಶಕ</i>	acaaategat	getenacttg	gttegetgge	egitgræckejt	9000
97750009649	tettgragaa	asocagocta	888888888888	aagagcagac	tcatgcarsa	9060
otywysttee	tgcaacetaa	gtteagtaat	caggogotet.	&caactggct.	grytgytopa	9730
tugtorgema	tttacttooz	gttctatgat	ttaaoggtag	ctopctgttt	gatggcygaa	9180
alggartate	getgggagae	taaogatace	gcatczcycz	ttateaaacc	eggege etgg	9240
caygge acc c	atgooggttt	getegog gg t	gaoncuttaa	tectmatet	ç çcaca gatg	9300
gaagatgeee	acctgaaaca	ggutaaange	gtactggagg	tagazogtac	ogtitegetg	9360
googeegtet	atgoomaact.	асодсвадах	aaatttatoo	tgactcagga	aatagagaag	9420
ttogtgagta	aaggttcagg	candddacoddc	aaggecaata	atabggtggg	gtttagtæcc	9480
autaccasts	octototaga	agogtocatt	togitatota	cctogeacsi:	tageagegat	9540
tatoctgett	ctattgytaa	aaceogtegt	attaaacaga	toagrytter	cetgecagea	9600
ctgctaggac	cctatcagga	tgtgcaagca	attetgtett	асьдордава	egaatakgea	9660
ttggctgaaa	gttgcamate	attagoggtt	tet e abggga	tgaatgacag	oggtcegttc	9720
caactggatt	traergstigg	(Asphinotg	cognitogeag	gaatcaaaat	egatgaaggt	9780
acgstgemat.	tgagettece	aastgcaatt	agtaaagsag	acertands.	сдахмаадде	9840
АВЭСВАСААС	ccatgctgga	gagtetgaae	gacatoatto	tgcstattmy	ctacaccatt	9900
ogcomatasc	gattttaatt	aagtgctaaa	acagyoccet	aagcggggc	tocaaggagt	9960
ctttcatgts	aaattcacaa	gatttcagta	ttaragaact	atcattgood	<u>ажаддардалд</u>	10020
					atgycogogo	
tgtetetgee	gttgectate	tetgeeggge	grogetatge	tergteacte	9ccttamact	10140
					gütatgadda	
tocycegoog	cacccatttt	ggegtteeae	attatgatga	aacogatacc	tttetgggge	10260
cagatggcgs	ggtectggtg	gtayrggatc	aztoongoga	Cyastogaca	ttacagggta	10320
teaacttagg	caucecettt	acceptiacrog	galac ogt te	costotypag	agtcatttca	10380
geogattgga	atattggcaa	оссавдески	CHCCCSagac	авсодуська	acagattttt	10440
فاعدموا 1950	(taggoorageft	ggacsagtec	stitectgyg	tasetoxoca	CONSCOORING	10500
lioag cas cee	gtoægacatc	acticaaacag	cacaatgytt	getagaagec	tetgtgtese	10560
cocatggtga	acesatitat	tatçeatetc	digocarday	taacacoggt	tg cgaa getg	10620

abyaaattac	tetecation	caggeogeog	ograaogtta	tetacscaca	gtgtattacg	10680
GOARCAGGEC	воссвосава	acgttacocg	gtctggatgg	ಧಾರ್ಷದ್ವರ್ಥ (೧೯೩೪)	CCSCSSGCSQ	10740
ac:Cggttatt	ctatctggts	tttgattang	dadersodase	taacaacetg	прододожен	10B00
cagnatitite	gactacaggt	agetggettt.	gt.ogccagga	rogtttttee	ogttatgaat	10860
atggttttga	gattogtacc	egeegettat	gcogtcaggt	stigatykat	nactatetge	10920
aagetetgga	tagogagata	<i>DOMESTICA</i>	acggaccaac	getggtttca	ത്യാലുപ്പാര	10980
teaattatga	стражения	atugocagoa	cyclygtatt	cyttegtoga	gtaggocacg	11040
agodágacgg	tactgrogic	accettgeogo	cattagaatt	ggeqtateag	qatttttcac	11100
rgotacaltaa	oga tag at g g	caatogatga	atgtgetggge	aaacttcaat	gecatteage	11160
go bggcsac ti	agtigalista	8689 30 9088	9atteecom	tobgotatat	caagetaaag	11220
gogootggtg	gtaccyctcc	goscasogkt	Ulgodanaat	togctoegat	gregteartt	11260
gggaaasat	gcaacctttg	toggttatec	ettockteca	aagtaatgoo	togetggtgg	11340
atatoaatgy	ඉසිප්පසුදිපත්වෙන්	cttgactggg	ttatcaccyg	ac re ggatta	ogggatate	11400
atagtoagoa	tecagatgge	agttggarac	gttttaccc;	gotomacgot	ctgccagtgg	11460
aatatactco	tocaegogeg	caactegoog	attteatggg	agetggaett	totgatttag	11520
tactgategg	COCTANGAGE	gtaogtttat	atgccaatac	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	tttgccaaag	11580
goossagatgt	agtgcaatcc	ggtgatatca	cartgecagt	පයයට්ට්ට ර්ථර්ර	gateogtgta	11640
soppagninge	atttagtgat	gtattgggtt	coggtcagge	acatotygtt	ධ්පපලි ාධ්මේරුවි	מסלוב
ogsctæsyt	czectycty	wtestotgg	dicredined	ttttggtcaa	ccaattactc	11760
tteegygatt	tagecsaces	gesgogeogt	Ctantoctge	tcaagittai	ctggccgate	11820
tagatggcog	oggeoogaet	gatotgattt	atgittacec	agategtetg	gatatettee	11880
tgaataaag	්ර්රියම්ටේ ට	thogrageac	cagtaactet	cocctteres	geeggagtge	11940
gttttgatca	tacctgtcag	ttaceagtgg	ompatigitaca	agggttagge	gtegecagee	12000
tgatattaag	tgtgoogcat	atgacteroc	<u>Btcactggog</u>	t.tg cg atetg	acceacacaa	12060
eecogtggtt.	acticagtigaa	а спас наса	atatgggggc	tcatcamack;	ctgogttacc	12120
ytagttoggo	caagtictagg	otggat gaaa	sedeca odd c	actggatgcc	ggacaaatac	12180
cagtttgtts	tetaceette	coggtacaca	ccctatggca	aacqqaaata	gaggatgaaa	12240
t.cagggggaat	CB&B&CTAGEC	a caa tactae	gttatgcacs	tggegeetgg	gatyganging	02300
agogegaal;L	togoggatti	gyttatgttg	aaragaaaga	cagocatosa	срайсханна	DARSC
grægtgeger	agaetycaca	concetecae	tgacccaagg	caacgogoot	gaactcacet.	12420
cacorgoget	gacorraagge	eacoxxtocau	aautxweace	acctgcpatg	ассоваадся	12480
acgegeetga	acteacates	coopeyetas	⊂ෆ්ටෙඵඉඉ¢සෙ	ододосождая	ttcacatcac	12540
cogregategan	ccaaggcaat	<u>ўфросадав</u> с	tracaccacc	tpegatgace	aaaaactggt	12600

atgecacogg aatocccotg etagatamos cuttatogae ogzagtattgg catggtgatm 12660
acceangethit throughout theorethic greatestic cannellatic 12720
tyckcacacc ggaaaatgat aacagtcogt actggctaaa cryggcactg &aacggbcaac 12780
tgctacgcag tgaactgtac ggoggggatg googtacaca ggaaaaaatt coctacacoug 12840
toactgaatt togcccacag gtaogtoggt tacagcatac cgatagcoga taccttgtgc 12900
tttggtomte tgtmgttgmm øgeogemmet mtomttmogn mogtatoger mgegetoete: 12960
estgrapota amagattacg cumicrescy slutaliting tomocogeta amacegniti 13020
ogytacogta boueczongo negometnyg campungton ytatoctymi acyttycoty 13080
ataagttatt testaacae talgalgace agcaacacaa attaoggete acetateaac 13140
agituagitig gnatkis(city accessests complicity at gitinggatis coggetagia 13200
coopeagegs tatetitiget tatagegetg aacstytexx: tactogtogt ctamaicteg 13260
aaateetaaa tystaaaaat agtetgatto oygagaataa seedogtosa taeeteggee 13320
agezaazase egittataee gaegggesaa atgezaegee ategesseedg eezaeaegse 13380
assogutgat tyeottoso: gagacaecag tattizatos atomacação teagogetig 13440
altegenetat etesteleet exatteteka epaceetega acaageegen taccageaaa 13500
cayattatet attermyege artggagaag atamagtetg ggengetegt egtpgetatm 13560
ctgattacgg cacagoogsa caghfinbggc ggccgcsaan acagagosar Actonacton 13620
ogguraasat caegeteaet togguatoraa artathoogt ogtososcas accopymate 13680
egyptogaet gacaacetca gecagatatg attggegtti: tebgaccece gttcaacters 13740
ogyatatona ognosatosą escritares egetygatyk actygatosa ecsatescac 1,3800
typpottttg ggypaccgaa aacggtaaga tgactggtta ttottoacog gazaaastal. 13860
cyttitetee acceletgat yttgseyoog eyattaagit sacunogoes atcertgiag 13520
cacagigica gyietangea ocegsaagei gysigemeat attamagama americanta 13980
асструкацы дтаждацору азадэдттак atascamxxy ажтохтогос давдагуудж; 14040
gostotytec cotegoticae egocyntygy taasaayoos saytycsyte accesgoesa 14100
topatotyte anaegycagt eccepttiae coeffcatay outcoeffty actaegyate 14160
gttatgaccg cyatettaag caacagatte gtesacwagt wytatteagt galggetttg 14220
geogettaet graagestet geacqueaty жерохухода agretgycka cytanocuan 14280
acqqcqctct ggtgaceaea акуунадары коммалодод ctgggcqgtt моддуасрса 14340
ctgsatatgs castaaggga casccyster; gescetatea secctattic etemaogmet 14400
gymaataegt cagtaatgac agtge $cogg$ c gymcagaaga agoctatgca gymcaccatg 14460
Extatgator cattggtoga gasatosagg Exactacego aasaaggotgg Ebucgtogaa 14520

octtettican; tootteetti acteticaate aagateessa teacacagut acteagytes 14580
ayytaaaysa gasagaatgi aaagaaggia aagaaggisa waxigisatt tgatcaatcc 14640
ogcomggttg sagggoggga aacataacat satatagagg tganaogtgt cattcataat 14700
gungknagek actomactia tgayttggft gatcettggt titattgogg cotggggogg 14760
ettagtaggg tacctcattg atatacases casacoalgt montggaatt ggatcaeogt 14820
actetytema etestiatet eetyttitae egytatelika aangasetye tyagittiga 14880
aagoggegge agoooctatz texettttge gattgeoggg etatttgges eczogggaag 14940
ttotggattg aactggatet ggogtogret ttttatgeat tatogegatg atggaggaaa 15000
graataagge attoccacty organiaaan estetytete oggonistiaa arogggaast 1.5060
taootaetae aastattgta agaaaaegaa tatatagaaa aastazsatg eagataasaa 1,5120
otpogattpc agraecogatg acacacaacg coccaacaan gaggtamatc atg aaa 15176 Met lys 1
asc atm get cot eas cit tab cas and not cot gio gio eas atc tac 15224 asm lie Amp Pro Lys Leu Tyr Gin Lym Thr Pro Val Val Amn lie Tyr 5 10 15
gst aac ogs ggt cta sog ato ogt aac ate gsc tit cac ogt acc acc 15272 Asp Asm Ary Gly Leu Thr Ile Arg Asm Ile Asp Rha His Arg Thr Thr 20 25 30
gca aac ggc gat acc gat atc cgt akt act cgc cat caa tat gac bcc 15320 Ala Aan Gly Asp The Asp Ile Arg 1le Thr Arg His Gln Tyr Asp Ser 35 40 45 50
ctt ggg can cha agu caa agu acc gat eeg egt cha tat gea gec aaa — 15368 Leu Gly His Leu Ser Gln Ser Thr Asp Pro Arg Leu Tyr Glu Ala Lys 55 — 60 — 65
can man tot amo tit etc igg cag tat gat itg mor ggt mm: att itg 15416 Gin lys ser Asn Pho Leu Trp Gin Tyr Amp Leu Thr Gly Amn Ile Leu 70 75 80
tgt aca gaz age gte gat get ggt ege act gte ace ttg ast gat att 15464 Cys Thr Glu Ser Val Asp Ala Gly Arg Thr Val Thr Lou Asm Asp Ile 85 90 95
gam ago ogt oom ota otg ama gta aut gom aca ggt gto ata oas aux 15512 Glu Gly Arg Pro ien Leu Thr Val Thr Ala Thr Gly Val Ilo Glu Thr 100 105 110
ega caa tat gas ang tot kee ota eee ggt egt etg tig tig git ace 15560 Arg Gln Tyr Glu Thr Ser Ser Len Pin Gly Arg Leu Leu Ser Val Thr 115 120 125 130
gaa caa ats cms gea exe ace too ogt atc ace gaa ogd etg att tgg 15608 Glu Glu He Pro Glu Bys Thr Ser Arg He Thr Glu Arg Leu He Trp 135 140 145
get gge aat age gea gea geg ama ame ent aat ett gee age emg tye 15656 Ala Gly Asm Ser Glu Ala Glu Lys Asm His Asm Leu Alm Ser Glu Cys 150 155 160
gtg oge cac tat gad ang gog gga gtd zon oga tta gag egi: titg toa 15704



Val	МÜ	618 1 6 5	ТУT	Авр	Thr	Ala	Gly 170	Maj	Thr	AYJ	Leu	Glu)75	Ser-	Leu	Ser	
							caa Cln									15752
							gat Asp									15800
ctg Lei	get Alæ	gat Asp	gac Asp	atc Ile 215	tac Tyr	aca Thr	acc Thr	ctg L e n	age 800 220	gcc Ala	ctil. Phe	9at Aep	goc Ala	acc Tim 225	eja General de la compansión de la compa	15848
got Ala	tta Leu	ete Leu	act Th r 230	G)n cag	acc Thr	eet qaA	g¢g Ala	ова Lyв 235	Gly 993	aac Asti	att Ilo	Cag Cag	вдд Ахд 24 0	cta Leu	noc Tim	15896
tat. Iyr	gat Asp	gtg Val 245	gee Ala	Gly Gly	CDD CBG	cta Leu	дес Ав п 250	ejà asa	agc Ser	tag Tap	tta Lep	acc Thr 255	tte Leu	esa Lyb	Pab Bac	15944
Gln Ges	eeg Pro 260	ðĮ≀i δæ∋	ajn cas	gtg Val	att Ile	atc 11æ 265	Arg Arg	tee Sor	ctg Leu	ácc Thi	tat Tyr 270	Ser.	N) θ ge¢	gee Ala	gga Gly	1,5992
caa Gln 275	aaa Lys	tta Leu	Arg Ogc	gag Glu	ള്ളു Gl v 280	cac Bis	Oly Ogc	aat Asn	GJY GJY	gtt Val 285	atc Ile	acc Thr	ರು ಚಿತ್ರಾ	tac Tyr	agt Ser 290	16040
tat Tyr	Glu goa	ccg Pro	gaa Glu	acc Thr 295	cæa Gln	cag Glu	ctt Ieu	atc Ile	ggt Gly 300	acc The	aaa Lys	ecc Thr	cac Ris	cgt. Ang 305	eeg Pro	1608A
tma Sec	gat Asp	ger: Alæ	ал в LVS 310	V∋1	ttg Leu	caa Gln	дас Авр	cta Leu 315	ogt Arg	tet Tyr	Glu Glu	tat Tyr	9560 Asp 320	cog Pro	gta Val	1 6 136
egy age	aat Asn	gte Val. 325	Ę	agt. S c r	atc ell	egt Arg	ast Azn 330	9ac Aep	gca Ala	gaa Glu	gec Ala	acc The 335	oge Arg	ido Phe	tog Trp	16184
cac Ris	aat Asn 340	ĠΪ'n	aaa Lys	gtg Val	geg Ala	erg Pro 345	gea Glu	aac Aen	act Thr	tat Tyr	acc Thr 350	tac Tyr	gac Aap	toc Ser	ttg Leu	16232
tat Tyr 355	Glia	ett Leu	atc Ile	Ser agc	ಥದಾ ಸ್ಕೃತಿ 360	acc Thr	GJA ggg	ege Arg	Gìu Gìu	atg Met 365	γŀa	aat Ago	ats Síl	ggt: Gly	cag Gln 370	16280
caa Gln	egt. Ser	ABN ABN	caa Gln	ett Leu 375	Pixi	tec Ser	ete Leu	acc Thr	cta Leu 370	Pro	bot Ser	gat. Asg)	aac Aan	aac Asn 385	ecc Thr	16326
tac Tyr	acc Thr	ABT BBT	tat Tyr 390	Tìm	cgt Arg	act Thr	tal: Tyr	act Mr 395	TYE	gac Asp	egt. Arg	ejà ôôc	990 Gly 400	Asn	ttg Iæs	16376
act Thr	F\$/3 688	ato 11e 405	Glr	cac His	agt Ser	toa Ser	cog Pro 410	gog Ale	acg Thr	caa caa	aac Aan	ക്കുറ Asn 415	Tyr	acc Thr	aca Thr	16424
aan Asn	ate Ile	acy Thu	yej Vej	tet Ser	aac Asn	ogg Arg	agc Ser	ast Asn	cyc Arg	gco. Ala	gta Vel	ete Lou	age Ser	act. Thr	ren tañ	36472

WO 99/42589

- 32 -

	6 20					125					430					
acc Thr 435	gaa Clu	gat Asp	ceg Pxv	geg Ala	caa Gln 44 0	gta Val	gat Xe p	get Ala	tta Leu	ttt Phe 445	gat Asp	gca Ala	ggc Cly	gyy gga	cet Ris 450	16520
								aac Asn								26568
								aaa Lys; 475								15616
								agt Ser								16664
								aac Asn								1.673.2
								cta Leu								36760
								gta Val								16808
								ggt Gly 555								16856
			AIT					Asn					Ser		ett Leu	16904
		λερ					$\mathbf{H}_{\mathbf{G}}$					Glu			ece Pro	16952
	Gly					TIE		aca Thr			Arg				agt Sar 610	17000
tat Tyr	aaa Lys	acc The	ate Ile	Arg 615	Tyr	tca Ser	ggt	:aaa 'Lys	g∋g Glu 620	. Arg	gat Asp	goo Ala	acc Thr	gga Gly 625	cta Leu	1704A
				Tyr					Pro					TI.	tta Leu	17096
agt Ser	gee Ala	gat Asg 649	Pro	gca Ala	igga Gly	aca Thr	gta Val 650	Asp	GJY GGG	t.t <u>o</u> Leu	e dere Ren	tta Ien 659	Tyr	cyg Arg	atg Met	17144
gtz Val	аду Агу 660	ÀЮ	: 821, 1. <i>83</i> 7,	Pro	gtt Val	act Thr 665	. Ter	ctt Leu	gat Aeç	ect Pro	. qat Аур 670	Gly	tle Leu	atg Net	Pro	17192
	Ile					: Ala					Aen:				уні. Аэр 690	17240

. 1

um gog oct tog o Ser Ala Pro Ser P 6			17288
eng ere gta god e Pro Pro Val Ala P 710			17336
cas tos act acm t Glm Ser Thr Thr T 725			17384
tog ggm toa to: H Ser Gly Ser Ser 1 740			17432
act not gas ata t Thr Pro Glu Ile 8 755			17490
get all tes aes a Ala Ile Sar Thr A 7			17528
get aat aga too t Asp Asn Arg Ser P 790			17576
ass got tog oct o Lys Ala Trp Thr 9 805			17624
gto thi all ggt o Val Phe Ile Gly G 820			17672
Ang mat ata aac a Lyn Ann Ile Ann T 835			17720
act tac ata ana t Tur Tyr lle lys T			17768
goa att aat wet g Ale Ile Aen Thr G 670			17816
ges wil aat atg g Glu Ile Aso Met A 885	. Phe 'Thr lle Amp		17864
aat oos ots ors a Asn Pro Leu Pro A 300		Pro Ser Leu Leu	17912
ctit gad ama dea q Leu Amp Thr Pro (915			17960
gga cong gta aat q Gly Pro Val Asn &			18008

aan aat gta ega ket tat aag aga taa egamaaatta atattettta 🦠

- 34 .

PCT/EP99/01015

18055

lys Asn Val Lys Pro Tyr Lys Arg tetaettita atageeetet tgaacttaea eteaaggaga ggaazeeaaa taagaaacea 18115 telttaetan exacecatga aagsatatti atticatgge tigattaett itaacattea 18175 otattaanu sitaaaacaa tatotaacca attaaaataa caatacotta titatoatat 18235 Leassietra aatoagaaat taatgaatit aagggttoit tatzittati toigagagoa 18295 taggeseset accttacoga taggesetass estastess astecagaas tactatatt 18355 testeaatat gogragusta gografitos tigagagica timaacitat egogragueco 18415 octitizeda gatoczatet aftaglasas traggosorg trastazope tasatitigg 18475 tattosgggu gataattosa tygoatsaaa Utattooatt gittitaaaaa agoactatta 18535 tgotgaacna aaggaacaut agal'altatt toatosgogli geotittotgg titotaacata 18595. teatoggata cegcasgege cagestitgs teagescest chatcolgat gatatoates 18655 ttaluttuust aasattosot optossataa astottatat ossottiost Laasonaan 18715 aatamograf ottgoottat taaatoatoa attagaaaat tgttgattat amaastatog 18775 ogataatgat eacyttycau cyctotitti acyaccytay etaittiait aacataiici 18835 ecactigige castascos; titigiciets titigatasti tataattiet acgaraatte 18895 caattottet caacttteag gateetttea taacaeggea geaactettg atatagtgee 18955 tttesstatt stytesystt gytttttoss gyteytegat oseacaattg acassecaca 19015 omotyttoca ettektatan eagooteeta astegasaag egytamtaca aagogtatee 19075 geogetäätig ligaatgadin läikelääägel galteenakaa aataotttaag ttyotttyssa 19135. caasatates tescatacee tettytiite atteeagasa tagaatutta accatagaan 19195 atgaceacys tythiciant thyceitoit thecallegy acatgoptia atgymcatty 19255 matticects catcastigt tastatibat timatectic coccatant; stamming 19315 tataectteg tteattatti: cttgatattg atvatggtes gttttoctca ataoctacag 19375 aagtagatal 148ffklakút toragtaato tatogtitigg ogaoggaggi egattettee 19435 attgggetat tomaccomit ogoogoolii ollattaatt acegtgattt ttggcatttt 19495 ggUttrater aactteggüt betaggtget bulkesttta gescoogstg ttaactteau 19555 cotaaaggga tarstaceaa etteacettg taagaatatt etgtttggte tacetteaac 19615 gactiticama atggggtama taacceggget asaatcaate gtatecaatg catcaatitic 19675 gotgatatit yteogggeig estesityat aasigegati aaaloggiiy chgaalageg 19735.

astagratot ticactagat garggarate ggtataarte actgarzean aggrixggix; 19795 astettorae tiacatogae egecacrati алааддіаді titigecigaa agtazerggi 19855 titoggatos getilliseal erzgargian kenginskan gitogiaect талааддода 19915

PCT/RP99/01015

catattopaa totaaaogat ahkkaaggos akettitigag atalacadag oggatacatg 19975 eggetgtgtg talbtagglig cyanichtic iacagtamic cactgatict cluteggagy 20035 apagagegge teatttgogt (Ageacages tgatasteaa atsacagata agacagataa 20095) quanticing etaittatca teglaselit tentesello clacagogit atotocatot 20155 gigteenatt compatette eigitimiste attimgement mamitageta egetemisee 20215 attactical alifection atgastogas astegacite tigagiging (talcastit 20275 tocogramy ataticasco cactogorgo ettlettati agicacegig sociingens 20335 tititogitic atomagetta ggettaaaaa taattitoos titagetoot ggagttaacg 20395 byngtlyana aggacyczti titaatacti czecttytaz gaataticty theygynyau 20455 ottownogan iticaasaca ggytaaataa ceggetasa atcaatogia ilomatetog 20515. agatitigi aataticato iggaciaigo catigataga igogatiaaa coggilgoig 20575 aetarggaat agcatottto soosgatggo typostoagt ataactrano gatecsassyy 20635. ecogyttaat tittooatita entegioeee elooatiaaa agytagitti getigaasai 20695 amongyttty tegatoggod thomolistos geogeageon attalogyto gyomotttam 20755 azgropacat attegratice seaccetact caegocate ettigatate tatteteege 20815 atacabgtgg Ulxiggtatat htmggcgdta cocchictae cybyatocat tgattiticit 20875. ₹aggagggga aaggggctca tttgggtcag cacagcctga tAttamaatc actgacampa 20935 caaataagta ttititaaca titatcatgg taagtitico toaattoous cagcattato 20995 operatazata teetytesag satagoytte attystttey tescessaga sacsagatay 21055 todomatori: ettacomosg stacemaers cogottabge ogtgagtast agtgagtiga 21115 gogaczagoga taczączagtą categogate aattagteee titigzataaa gogzaczegas 21175 tttgsaacut cogtcatecc gtccstatta cygascttag attatgatta ttaaatcacc 21235 acceasiggs asgessett ticatititi astitacgas gastgaatti qiaaqasaqt 21295 gttakaaant taatagaaat taatiteetg itaalotant gaaggatgaa attataaaaa 21355 teacccattt ctcagggaca accatccaca atatétages coactogtoc tcacttaatt 21615 topbytosey agtagaasta topbyatyse tosyboyaty acataczyca atytoattyg 21075 tattgegect eccescigit taateesttt cittigict taatgegegeg etecaegige 21535 ttoactatti dagoactate gataaataag attoomaant agogomatat ettecaqqae 21595 toataattul aigtalaaco attgyttäää tapgalootg tytääcägga tiatgaaakog 21655. ttatttatet casatetate sattattita tatatagitt caragicars cimp(a)ci: 2171,5 ggtacettes tasceauty occioectyc getzeettet gatascaaca getacaetas 21775 ctataccogn gmmtataatt atgaccytyn gaaattoog cytagttoac cygecacyca 21835.



Mat Ile low bys Gly Il Asn Met Asn Ser Pro Val Lyu 960 965

gay Glu	eta Ile 970	izit. Pro	gat. Asp	val Val	tta Leu	ава Lyrs 975	alu: Ile	വു വു	l.g(; Cye	ggt Gly	ttt Phe 980	cag Gl n	tgt Cys	ctg Leu	aca Thr	23954
gac Asp 985	att Ile	age Sar	cac Big	agc Str	tet Ser 990	tti. Pha	aac Aso	gaa Glu	lut. ₽he	сас Иів 995	CJI) CSG	CJV C95	Qta Val	.5cr	gaa G2u 1000	23902
cac His	cte Iæi	tee Ser	Tερ	tee Ser 1005	goa Glu	gca Als	cac Ris	\mathbf{A}_{T}	tta Leu 1010	tat. 'Iyr	cet. His	gat Aep	Ala	con Gln 1015	cag G)n	23950
gee Ala	caa Cln	1395	gat Asp 1 0 20	aat <i>Asa</i> i	Arg Arg	ctq Lav	TyT	922 Glu 1025	gog Ala	egt. Arg	atı, Ile	Leu	апа <i>Ly</i> s 1030	oge oge	acç Tur	23998
aat Awn	Pro	caa ඩො 1035	tta Leu	caa Glin	aat Aan	Me Det	gta Val LO40	cat Nis	ett Lau	gesc Ala	ile	gta Val 1045	goç Goç	oct Pro	aat Asn	24046
Ala	ges Glu 1050	etg Leu	a ta Ilc	GJ√ GGC	J)AL	ർഗ്ര Aen L055	aac Aen	Cara Cara	ttt Øve	Ser	060 ცეგ მელ	agg Arg	gex: Ala	øgt Sen	caa Gln	24094
tat Tyv 1069	Val	gon gon	eeg Pro	$g_{\mathbf{I}\mathbf{y}}$	acc Thr 1070	gtt Val	tne Ser	tuxo Serr	Met	tto Blue L075	toc Serj	ecc Pro	gee Ala	Ala	tat Tyr 1080	24142
ttg ieu	act Thr	<u>ејп</u> өэд	Leu	tat Tyr 1085	ogt Arg	gaa Glu	gca Ala	ķχŋ	eat Aen 1090	tta Leu	lgs cac	gcc Ale	Ser	gst Asp 1095	top Ser	24190
g(.l: Val	lat 'lyr	AU	ctg Leu L100	gat. Asp	act Thr	yry cac	λrg	eca Pxo L105	gat Asp	ete Leu	aaa Lys	Ser	atg Met L110	Vja Ösğ	ete Izu	2423R
agt Ser	GIn	caa Gln i115	aat Asn	Met Met	gat. Asp	acy Tiv	gaa (3) ເລ [120	ett Leu	tee Ser	act Thr	Leu	tot Ser 1125	tta Leu	toc Ser	aat Mgo	24286
CLA	cts Leu 130	tta Leu	Ltg Leu	9aa Glu	Ser	att Ile L135	aaa Lys	act Thr	gag Glu	Ser	аа9 Lys (140	ctg Leu	yest) Yat	aat Aun	tat Tyr	24334
act Tur 1149	Clu	gtg Val	atg Mct	Glu	atg Met 1150	oto Leu	tee Sen	get Nia	Ff 100	ogt Arg 155	eeu Pro	loc Ser	Gly	Ma	acg Thr 1160	24382
eet Pro	tat Tyr	cec Gia	AST	get. Ala 1165	tac Tys	gaa Glu	aat Asn	Val	egt Arc 1170	TÅS Sæg	glit Val	abo Ile	Glat	cta Leu 175	caa Gln	24430
gat Aqp	ect Pro	GTA	ett leu 180	œg Glu	cas Gln	tla Leu	Asn	got Ala 1185	tex Seri	eca Sm)	gee Ala	He	9000 Ala 1190	CJ À 869	otg Leu	24478
atg Met	ມາຣ	caa Gln 195	got Ala	toc Ser	cta I <i>m</i> i	الاس الاسا الاسا	99¢ Gly 1200	att. Ile	acc Asn	get Ala	Sear.	atc Ile 1205	tca Ser	eet Pro	geg Glu	24526
ttg Jæl	t.f.l: Phe	aat: Aeri	att. Ile	otg Leu	oicg Thư	gag Glu	gag Glu	att Lle	act. Thr	gaa Glu	Gly ggt	aat Asn	get M.a	geg Glu	oyaa Çlu	24574

1210	3215	12.	20	
ctt tat ang AAA Leu Tyr Lys Lys 1225	aat tiit ggli a Aso Yhe Gly A 1230	aat ato gea ong g Asn 11e Glu Pro A 1235	ct toe etg got atg la Ser Leu Ala Met 1240	24622
Fro Glu Tyr Leu	aga egt tæt t Arg Arg Tyr T 1245	tac ast tha agulg Tyr Aen Leu Ser A 1250	ජෝ ඉහස ඉහස අප්ප ලෙද කුර මටය මටය Less දිදැල 1255	24670
cag tit att ggt Gln The De Gly 1260	aaa god agd a Tays Ata Som A	aat tte gge caa c Asn Phe Gly Gln G 1265	8A gas tat agt aat In Glu Tyr Ser Asn 1270	24718
aac caa ete att Asm Glm Leu TJe 1275	The Pro Ile V	gte aae age aat g Val Asn Ser Asn A 280	at ggc acs gtc awg sp Gly 19m Val Lys 1205	24766
gka fat oga att Val Tyr Arg Ile 1290	acc aga gas t The Arg Glu 1 1295	tat acs are ask go lyr Thr Thr Ash A 130	oc aat waa gta gac la Asn Gln Vel Asp 00	24814
gty gag ctg tit Val Glu Leu Phe 1305	eee tae ggt g Pro Tyr Gly G 1310	gga gea æat tat c Ny GJu Aso Tyr G 1315	ag tia aan tac ooa In Lan Agn Tyr Lys 1320	24862
Phe Lye Amp Ser	nyt cag gat g Arg Gln Asp V 1325	gto too tat tta to Wal Ser Tyr Teu So 1330	co ato ega tie egi ev Ile lavs len Asm 1335	24910
gac aan aga gaa Asp Lys Arg Glu 1340	Leu lle Arg I	att gma gga geg e Ile Glo Gly Alm P 1345	ot cag gio aac atc to Gin Val Aen Ile 1350	24958
gas tat tes gas Olu Tyr Ser Glo 1355	His Ile Thr 1	its agt ava act g Leu Ser Thr Thr A 360	ut ate agt caa eel. sp 11e Ses Cln Pro 1365	25 00 6
tit gaa ate gge Phe Glu Ile Gly 1370	eta aca ega e Leu Thr Arg V 1375	gta tat ook tot a Wal Tyr Pro Sor 3 13	gt tal byg goa tat er Ser Trp Ala Tyr 80	25054
gea grot ges eas Ala Ala Ala Lye 1385	the acc att g Phe Thr Ile G 1390	gag goa tat aac c Rio Gio Tyr Asn G 1395	aa tan bot tto otg In Tyr Ser Fhe Len 1400	25102
leo Lys Leo Aen	ana got att c lys Ala Ile A l405	ngt eta tet egt g Arg Len Ser Arg A 1410	cy ach gaa tta tch la Thr Glu Leu Ser 1415	25150
cxx: ecx: ett ctg Pro Thr Ile Leu 1420	Glu Ser Ilo V	otg ogt agt gtt a Vel Arg Ser Vol A 1425	at cag coa chy gat so Gho Gho Leu Asp 1430	25198
ate ame gem gan He Asn Ala Gio 1435	Wal Leu Gly 1	aka gil tit otg m Lym Val Phe Lea T 440	ot man tat tat atg hr Fwys Tyr Tyr Met 1445	25246
csa cgt tat gct Cln Arg Tyr Ala 1450	att ast got g The Asm Ala G 1455	ysu net god eta a Slu Thr Ala Leu I 94:	ta ota tgo aat goa le 1an Cyc: Aun Alm 60	25294
ctt att toa coa Loo The Ser Ghn 1465	ngt toa tat q Arq Ser Tyr A 1470	gat aat daa ook a Yay Ash Gin Dro S 1475	gu twa ttt gat ege er Gln Phe Asp Arg 1480	25342

ntg tit aat acg oom tim olg ome gge ome law u.d. bet mee ggm gem. Lew Phe Per Thir Pro Lew Lew Aem Gly Glin Tyr Phe Ser Thir Gly Aep 1485 1490 1495	25390
gas gag att nut the aut oce out agt act gar out too out too out too of the Clu I've Asp Less Ase Pro Gly Ser The Gly Asp Trp Ang Lys Ser 1500 1505 1510	25438
yty off am dyf you tit mut ato gat gat mit too ofe too top ofg Val len Lye Ary Alm Phe Arm Ile Amp Amp Ile Ser Len Tyr Ary Len 1515 1520 1525	25486
ott ase all soc aac cat aat aat caa gel gga aag all aan aau aan; Leu Lys lle Tho Ash His Ash Ash Glo Asp Gly Lys fle Lys Ash Ash 1530 1540	25534
lis ent ast off for gat the fat ent ggg ean the only gow gew att len Ash Ash Lou Str Asp ion Tyr lie Gly Lyn Len Lou Ala Glu lie 1545 1550 1560	2558 2
cat caa tta acc att gat gaa ttg gat tta ttg ctg gil gro glg ggt His Glo lau Thr lle Asp Glu Leu Asp Leu Leu Leu Vol Ala Val Gly 1565 1570 1575	25630
gam 990 gas act. Ast tok tok got ato agt gut ann cas otg gog gos Glu Gly Glu Thr Ast Leu Ser Ala Ile Ser Asp Lys Glu Iku Ala Ala 1580 1590	2567н
ctg atc agu axx étc aat acc att acc gtc tog eta eag aca eag aag Leu lle Arg Lys Leu Asn Thr lle Thr Val Tro Leu Gln Thr Gln Lys 1595 1600 1605	25726
tgg sg!; gcg ttr ræa tha ttr gtt atg act tee acc age tat aac aaa TTP Ser Ala Phs Gln Leu Phe Val Het Thr Ser Thr Ser Tyr Api }gs; 1610 1615 1620	25774
acg ctg acg cct gas att aag aat ctg ctg gal acc go; tec cec ggt. The Leu The Pro Glu Lie Lym Asn Lou Leu Asp The Val Tyr His Gly 1625 1630 1635 1640	25822
tta caa ogc tit gat aaa gac aag gos aak ika okg oat git atg gog Lou Gin Gly Phe Asp bys Asp bys Ala Asm Leu Hes His Val Met Ala 1645 1650 1655	25870
eec tal att gog goe ace tla caa tla lea leg gas mal gto goe oat Pro Tyr Ile Ala Ala Thr Len Glm Len Ser Ser Glu Asn Val Ala His 1660 1665 1670	25918
Ser Val Leu Pro Ala Aep Lys Len Lys Pro Cly Asp Gly Ala Met 1675 1680 1895 1895 1895 1895 1895 1895	25 9 66
aca see gaa ama tte tgy gae tgg tty amb ach ach con oca gat The Ala Glu Lys Pho Trp Amp Trp Leu Amn Thr Glu Tyr Thr Pro Amp 1690 1695 1700	26014
2000 1100	
toa tog gaa gta tha gos ace meg gen oek ett gtt can tat tge can Sor Sor Chu Val Lou Ala Thr Gin Glu His Ile Val Chn Tyr Cys Gln 1705 1710 1735 1720	26062

asc gyr the ege etg til gig aca asa era gag atg til gge tog tog Asn Alm Pha Arg len Pha Val Thr Dys Pho Glo Ket Pha Gly Ser Ser 1740 1745 1750	26158
act gag ges gts eet geg est get ges ett tes etg ate atg etg seg Thr Glu als Val Pro Als Ris Asp Als Leu Ser Leu Ile Wet Leu Thr 1755 1760 1765	26206
ogt till goa gat tog git aat gog bia ggo gas ean god tot bor gia Ang Phe Ala Asp Trp Val Asm Ala Leu Gly Glu bys Ala Ser Ser Val 1770 1780	26254
cta gog goa itt gan got aac agi tia acg goa gaa caa the got gat Lou Ala Ala Rie Glu Ala Amu Ser Leu Thr Ala Glu Glin Leu Ala Amp 1785 1790 1795 1800	263D2
god atg aat ett gut got aat tig die itg daa ged agt ach daa ged Ale Met Asm Len Asp Ala Asm Len Len Cim Ala Ser Thr Gim Ale 1805 1810 1815	26350
caa aac cat can cat ctt coc coa gtg acg caa aaa aat gct; ttc tcc Gln Aan Rie Gln Rie Leu Pro Pro Val Thr Gln Lys Aan Ala Phe Ser 1820 1825 1830	26398
tột (gọ sas tot ato gọc act atr cho cảm tạp git aat git gia cam Cym Trp Thr Ser Ile Asp Thr Ile Leu Gla Trp Val Asa Val Ala Gla 1835 - 1840 - 1845	26446
cas tig sat gic gcc coa cag ggs gil tox gct tig gic ggg cig gst Gln Leu Asn Val Ala Pro Gln Gly Val Ser Ala Leu Val Gly Leu Asp 1850 1855 1860	26 494
tat att can tta aat can aan atc ooc acc tal gwo can ton gan agt Tyr lle Gin Leu Aen Gin Lys lie Pro Thr Tyr Ala Gin Trp Glu Sar	26542
1865 1870 1875 1880	
grt ggg gas ats tig am, gmc ggs tig ant tos cas cag grit gst sta Alo Gly Glu Ile Leu Thr Ala Gly Leu Aen Ser Glo Glu Ala Aep Ile 1885 1890 1895	26590
get ggg gas ats tig am, gee ggs tig ant tes cas cag get gat ats Alo Gly Glu His Leu Thr Ala Gly Leu Asm Ser Glo Glu Ala Asp The 1885 1890 1895 tts cac get tit tig gee ges tot oge sgt gee ges tit age suc tee Leu His Ala Pho Leu Asp Glu Ser Arg Ser Ala Ala Leu Ser Thr Tyr 1900 1905 1910	26590 26638
get gag gas ats tig and, god gas tig ant tes cas dag get gat ats Ala Gly Glu Ha Leu Yim Ala Gly Leu Aso Ser Glo Glo Ala Asp The 1885 1890 1895 tts dad get tit tig gad gas tot dag agt god gos tis agd and tag leu His Ala Pos Leu Asp Glu Ser Ang Ser Ala Ala Leu Ser Tim Tym	
get ggg gas ats tig am, gee ggs tig ant tes cas cag get gat ats Ale Gly Glu He Leu Thr Ale Gly Leu Am Ser Glo Glu Ale Amp He 1885 1890 1895 tts cac get tit tig gas gas test oge sgt gee ges tits age suc tac Leu His Ale Pho Leu Am Glu Ser Ang Ser Ale Ale Leu Ser Thr Tyr 1900 1905 1910 tat atc egt cas gic goo ang con gen gee gec ats ass age egt gat Tyr He Ang Glu Val Ale Lym Pro Ale Ale He Lym Ser Ang Asp	26638
get gag gas ats tig and, god gas tig and tos cas dag get gat ats Ala Gly Glu He Leu Thr Ala Gly Leu Aso Ser Glo Glu Ala Asp Ile 1885 1890 1895 the cad get tit tig gad gas tot ogd agt ged gos tis agd and the Leu His Ala Pho Leu Asp Glu Ser Ang Ser Ala Ala Leu Ser Thr Tyr 1900 1905 1910 tat atd ogt das gid god and odd gog god ats asa agd ogt gat Tyr He Ang Glu Vai Ala Lyb Pho Als Ala Ala He Lyb Ser Ang Asp 1915 1920 1925 ged tig the das the Ula dia att got and dag git tod god atd Ala Leu Tyr Glu Tyr Leu Leu Ile Asp Asp Glu Val Ser Ala Ala He	26638 26686
get gag gas ats tig and, god gas tig and tos cas dag get gat ats Ala Gly Glu He Leu Thr Ala Gly Leu Aso Ser Glo Glo Ala Asp The 1885 1890 1895 tis dad get tit tig gad gas tot ogd agt god gos tis agd and tad Leu His Ala Pho Leu Asp Glu Ser Ang Ser Ala Ala Leu Ser Thr Tyr 1900 1905 1910 tat atd ogt das gid god ang odd god god ats ass agd ogt gat Tyr He Arg Glo Val Ala Lys Pro Als Ala Ala He Lys Ser Ang Asp 1915 1920 1925 ged tig tad das tim the dat att got ast dag git tod god atd Asp Leu Tyr Glo Tyr Leu Leu Ile Asp Aso Glo Val Ser ala Ala He 1930 1935 1940 sas act acc ogg stC god gan god att god agd att das dug ted git Lys Thr Thr ang The Ala Glu Ala He Ala Sor The Glo Leu Tyr Val	26638 26686 26734
get ggg gas ats ttg and, goe ggs ttg ant tea cas cag get gat ats Als Gly Glu He Leu Ymr Als Gly Leu Asm Ser Glo Glu Als Asp The 1885 1890 1895 tts cae get ttt ttg gee gas tet oge agt gee ges tta age som tae Leu His Als Phe Leu Asp Glu Ser Ang Ser Als Als Leu Ser Thr Tyr 1900 1905 1910 tat atc egt cas gte gee ang ces geg gee sta ass age egt gat Tyr He Ang Glu Val Als Lys Pho Als Als Als He Lys Ser Ang Asp 1915 1920 1925 gee ttg tae cas ten U.s ets att gat ast tag gtt tee get gee atc Asp Leu Tyr Glu Tyr Leu Leu Yle Asp Asm Glu Val Ser Als Als The 1930 1935 1940 ass act see egg ste gee gas gee att gee age att cas cug tee gte Lys Thr Thr Ang The Als Glu Als He Als Ser The Glu Leu Tyr Val 1945 1950 1955 1960 ass. typ and Clu gas ant gts gas gas ast gee cat the ggg gtt atc Asp Ang Thr Leu Glo Ash Val Glu Glu Asm Als His Ser Gly Val The Asm Ang Thr Leu Glo Ash Val Glu Glu Asm Als His Ser Gly Val The	26638 26686 26734 26782

WO 99/42589 PCT/EP99/01015

The Typ Ale Gly Val See Gin Lew Val Tyr Tyr Pro Glu Aem Tyr Ile 1995 2000 2005	
gat one are significate ages that some has edgesty gam goal tha big Map Pro Thr Met Arg Ile Gly Glm Thr bys Met Met Asp Ala Leu Leu 2010 2015 2020	26974
caa too gto ago caa ago caa tha aat goo gat ach gto gaa guo goo Gin Ser Val Ser Gin Ser Gin Leu Aeo Ala Aeo Thr Val Gin Aeo Ala 2025 2030 2035 2040	27022
ttt atg tot tat ofg aca tog tit gag caa gig gek aat olk aan gik Mee Met Ser Tyr Leu Thr Ser Mee Glu Glu Val Ala Asu Leu lys Val 2045 2050 2055	27070
att agn gwg fat dad gaf, sat att aad aad get daa ggg dig add tal Ile Ser Ala Tyr Ris Amp Amn Ile Amn Amn Amp Glin Cly Lou Thr Tyr 2060 2065 2070	27118
till, and ggm oto agt gas act gat acc ggt gas tac tat tgg ogm agt. The lie Gly Lew Ser Glu Thr Amp Thr Gly Glu Tyr Tyr Trp Arg Ser 2075 2080 2085	27166
gte get cae agt aam tte age gme ggt mas tte gee get amt gee tig Val Asp His See Lys Phe Ser Asp Cly Lys Phe Ala Ala Asm Ala Trp 2090 2095 2100	27219
agt gna tog coc ace att got tgt oca att act oct tac oga age act Ser Glu Trp His Lys Ile Asp Cys Pro Ile Asp Pro Tyr Arg Ser Thr 2105 2110 2115 2120	27 262
ate ogt och ging alig bze ama toe oge tig tal eig ete igg tig gam The Arg Pro Val Met Tyr Lyx Ser Arg Leu Tyr Leu Leu Tro Leu Glu 2125 2130 2135	27310
rea sag gag atc act eas com acm gym amt ago amm gat ggo tat chi. Gin lys Clu Iko Thr lys Gin Thr Gly Asm Ser lys Asp Gly Tyr Gin 2140 2145 2150	27358
ace gag aca got tat ogt tat gag eta aaa tig geg eat ate egi tat Thr Glu Thr Amp Tyr Arg Tyr Glu Leu Lys Leu Aie Bis Lie Arg Tyr 2155 2160 2165	27406
god ggt aux; bgg aut aug oda atd aut tit gat gid aat gaa aaa ata Asp Gly Thr Trp Asm Thr Pro Ile Thr Pro Asp Val Ash Glu Lys lle 2170 2175 2180	27454
too asy ota 900 ota goa ama mmi mma gog oot gog oto tat tgt got Ser Lyo Leu Glu Leu Ala Lyo Ago Lyo Ala Pro Gly Jao Tyr Cys Ala 2185 - 2190 - 2195 - 2200	27502
ggl. taf. C25 ggt gam gat mog tig oig git aig tit tat mad daa daa Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Ash Gln Gln 2205 2210 2215	27550
got now ofto got ogt tat eam acc got tom atg cam ggg ctm tat atk: Any Thr Leu Any For Tyr bys Thr Alm Ser Met Gln Gly beu Tyr lle 2220 2225 2230	2759R
ttt god gut atg gaa tat aaa gat atg acd gat gga caa tac aza fxd. Phe Ala Aap Met Glu Tyr Tyr; Aap Mot Thr. Aap Gly Cln Tyr Tys Scr 2235 2240 2245	27646
TWI Arg Away Asso Ser Tym bys Glm Phe Asso Thr Asso Ser Val Arg Arg	27694

2250	2255	2260	
gtg WAL BAC CGC Val Asm Asm Arg 2265	tat goa gog Tyn Ala Glu 2270	gat tat gam att ooc tom tog gim am Amp Tyr Glu ile Pro Ser Ser Val Asn 2275 2280	27742
Ser Arg Lys Gly	tat gat tgg Tyr Asp Trp 1285	oga oot tot tot ote agt atg gta tat Gly Asp Tyr Tyr Leu Ser Met. Vol Tyr 2290 2295	27790
aac gga gat att Aga Gly Asp Ile 2300	Fro I'm 1le	agt tac ama goo ach ten agt gat tta Ser Tyr bys Ala Thr Ser Ser Asp Leu 2305 2310	2783B
ass also tat atc Lys lie Tyr Iie 2315	Ser Pro Igo	tta ags att att cat aat ggs tel gee Teu Arg Ile Ile His Asn Gly Tyr Glu 2325	2789G
ggg cay caa ogc Gly Gln Gln Arg 2330	aat caa tgc Asn Olo Cys 2335	aat ota atg aat aas tst ggk: waa ote Asn Lon Mot Asn Lya Tyr Gly lys ten 2340	<i>2</i> 7934
ggt gat aaa ttt Gly Asp Ly: Phe 2345	att gtt tat Ile Vai Tyr 2350	act age tig gga git aat dea aak aat Tim Ser Leu Gly Val Asn Pro Asn Asn 2355 2360	27982
Sor Ser Aso Lys	ctg atg ttt Leu Met Phe 2365	two two get that can that had gigh eat. Tyr Pro Val Tyr Gin Tyr Asn Gly Amn 2370 2375	28030
gtc agt ggg ott Val Ser Gly Leu 2380	Ser Gin Gly	aga tta eta tte ese egt gse ace aat Arg Leu Leu Phe Bis Arg Asp Thr Asn 2385 2390	20078
tat toa uni aas Tyr Ser Ser Lys 2395	Val Ciu Ala	tog att oot oga gen gga egt tet eta Trp Ile Pro Gly Ala Gly Ary Ser Lou 2405	28126
acu Ast cog aat Thr Asn Pro Ass 2410	get gee att Ala Ala Ile 2415	ggf gat gat hal get ave gne teg tta Gly Asp Asp Tyr Ala Thr Asp Ser Les 2420	28174
aac AAA oog AAt Asn Lys Pro Asn 2425	gat ott aag Asp Leu Lys 2430	can tan gin tat aig act gac agt aan Gln Tyr Val Tyr Met Thr Aip Sen Lys 2435 2440	26222
ggt act get acc Gly Thu Ala The	get gto toa Asp Va) Sor 2445	ggs cca gte gsC atc ast act goa att Cly Pro Val Asp fle Amn Thu Als fle 2450 2455	28270
tee oog gen maa Ser Pro Ala Lys 2460	:Val Gln Ve(i son gta ana gee ggt age aaa gee mee . Thy Val. Lys: Ala Gly Ser Lys Glu Glo 2465 2470	28318
ang litit acc gog Thir Phe Thir Ala 2475	. Авр шуз жеп	gic hoc all rag com too cot ago itt Nal Ser Ile Gim Pro Ser Pro Ser Pho 2480 2485	28366
gat mpa atg dat Ag) Glu Het As: 2490	tat caa ttt Tyr Glm Phe 2495	aat get ete gaa ata gat gge wa agl. Asn Ala Leu Glu lie Asp Gly Ser Ser 2500	2H414
otg ami tit act Leu Asn Pho Thu 2505	. asc sat toa [.] Asn Asn Scr 2510	ger agl act gat att oct tit der ged Alm Ser fle Amp Ile Thu Phe Thu Ala 2516 2520	28462

WO 99/42589 PCT/EP99/01015

ttt gea gag yat gge ogt amm otg ggt tat gan agt tte agt att oct Pre Ala Glu Asp Gly Arg bys ben Gly Tyr Glu Ser Phe Ser Ile Pro 2525 2530 2535	28510
att acc oge meg gbg egt ext get aat too etg acc etg ege cas mat lle Thr Arg Lys Val Ser Thr Amp Amp Ser Leu Thr Leu Arg His Asm 2540 2545 2550	28558
gas aat ggt gog twe Lef. Aty das tyg ggs gtd tat dyd att dyt dtt Ghn Asa Gly Ala Ghn Tyr Met Ghn Trp Gly Val Tyr Arg Ile Arg Les 2555 2560 2565	28606
eat ext. the fit get eye can the git geg ega goe act acc ggt aft Amo Thr Len Pic Alm Arg Glin Len Val Ala Arg Ala Thr Thr Gly Ile 2570 2580	2865/1
gat any aft cig agi atg gaa act cag aat att cag gaa com cag tia Amp Thr lle Lou Ser Mot Clu Thr Gin Amn Ile Gin Gin Pro Gin Leu 2585 2590 2595 2600	28702
Gly Lys Gly Phe Tyr Ala Thr Phe Val Ile Pro Pro Tyr Ash Pro Ser 2605 2610 2615	28750
act cut ggt gau gas cgt tgg Ltt aag cth tat atc aas cat glu ghi Thr His Gly Amp Glu Arg Trp Phe hys Leu Tyr Ile hys His Val Val 2620 2625 2630	28798
ga() ast aat toe out att ato tet toe ggt ceg ota aea get aca aat Amp Amn Asn Ser His IIa IIa Tyr Ser Gly Gln Lou Lys Amp Thu Amn 2635 2640 2645	26846
ata ago ace aog tia tit ato cot cit gai gai git coa tig aac caa Ile Ser Tur Tur Leu Phe Ile Pro Leu Aep Asp Val Pro Leu Aen Qin 2650 2655 2660	28894
get tod agu gru dag git tod atg acc tid mag man tod dum tom gat Asp Tyr Ser Ala Lys Val Tyr Met Thr Phe Lye Lye Ser Pro Ser Asp 2665 2670 2675 2680	28942
ggt acc tog figg ggs oct ose tit gut aga gat gat asa gga ata gta Gly Thr Trp Trp Gly Pro His Phe Val Arg Amp Amp Lym Gly Ile Val 2685 2690 2695	28990
aca ata aac CCL aaa too act tig acc cac tit gag ago gio aat gio Thr lle Amn Pro Lys Ser lle Leu Thr His Phe Glu Ser Val Ash Val 2700 2705 2710	29038
etg pat aat Wil agt age gom een atg got tte ege gge got ome oge Leu Asn Asn Ile Ser Ser Glu Pro Met Amp Phe Ser Gly Ale Amn Ser 2715 2720 2725	2908 6
ete tat itt igg ges etg ite tae tai ace etg afg etg git gee eaa Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro Mei Leu Val Als Cin 2730 2735 2740	29134
egt tig tig eat eag caa aac tit gat gas geg aac ege igg eig aas Arg Leo Leo Hee Gio Clin Aso Phe Asp Glu Ale Aso Arg Trp Leo Lys 2745 2750 2755 2760	29182
tat sto top ago one too gos tat att oft one coe cas att cas aat Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val Sis Gly Gim Ile Glim Asm 2765 2770 2775	29230

tst caa tog aac gic ogc oog tim tig gam gut doe agt tog aac agt. 2 Tyr Gin Trp Asn Val. Arg Pro Leu Leu Giu Amp Thr Ser Trp Asn Ser 2780 2785 2790	2927ê
gat cet tig gat tee git git eet gae geg gin geg eag eac gat eeg Agu Pro leu Asp Ser Val Asp Pro Asp Ala Val Ala Glu His Asp Pro 2795 2800 2805	29326
atg one tat aam git toa ace tit atg oge ace off gat ofg tig atc 1 Met Nie Tyr Lys Val Ser Thr She Wet Arg Thr Leu Asg Leu Leu (1): 2810 2815 2820	29374
gog dyd gge yad dai got tad dge dae thig gay dyd gat acg ott oad 1 Ala Arg Gly Asp His Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Aso 2825 2830 2835 2840	29422
gse gog aag atg tgg tat atg caa gog ctg cat ctg tte ggc gat and 2 Glu Ala Lys Met Trp Tyr Met Gla Ala leu His Leu Heu Gly Asp Lys: 2845 2850 2855	29470
out hat dig dog dig agt add ace ligg ant gau dos ogs dig gad ama ? Pro Tyr Leu Pro Leu Ser Thr Thr Trp Amn Amp Pro Arg Leu Amp Jys 2860 2865 2870	29516
god grg gat att act acc daa agt got dat tod agd top ata god got 2 Als Ala Asp Ile Thr Thr Gin Ser Ala His Ser Ser Ser Ile Val Ala 2875 2880 2885	29566
ttg cap cag agt aca tog gog oft tha toa ttg ogc agc god aat acc Leu Arg Gin Ser Thr Pro Ala Lou Lau Ser Leu Arg Ser Ala Aen Thr 2890 2895 2900	29614
etg ace gut ete tte etg eeg cam ate aat gam gty atg atg aat tae 1 Lau Thr Asp Leu Phe Leu Pro Glo Ilo Aso Glu Vol Mat Met Aso Tyr 2905 2910 2915 2920	29662
tigg caa acs tia gmi cag aga gita tac and etg ege cac aac etc tet. In Gir Thr Leu Ala Gir Arg Val Tyr Aen Leu Arg His Asn Leu Ser 2925 2930 2935	29710
ate gae ggt cag teg tta tat etg ees ate tat ger; acs cog gog gae : The Amp Gly Glin Pan Ama Tyx Lond Pro The Tyr Als Thr Pro Als Amp 2945 2950	29758
oog aaa gog tta ete age goe get git goe act tel: caa ggt gga gge ? Pro lys Ala Leu Leu Ser Ala Ala Val Ala Thr Ser Gln Gly Gly Gly 2955 2960 2965	29806
awy old odg gag tom tit atg the etg igg egt tie eng ean alg etg : Lys Leu Pro Glu Ser Phe Met Ser Lou Trp arg Phe Pro His Met Leu 2970 2975 2980	29854
gas ast get ege age atg gitt age eag oir acc cas the gge (cc acg Glo Aso Ala Arg Sor Wei Val Ser Glo Leu Tho Glo Phe Gly Ser Tho 2985 2990 2995 3000	29902
tta caa aat att atd gaa deb dag gad god god god dog etd dat gog tta : Lou Chn Asu Ile Ile Chu Arg Ghn Asp Ala Glu Ala Leu Asu Ala Lou 3005 3010 3015	29950
tta caa est caq god gos geg dig alta lig act wad otg agt att cas : Ieu Gio Ash Gin Ala Ala Giu Leu Tie Leu Thr Ash Leu Ser Ile Gin 3020 - 3025 - 3030	29998
gae ann aon ain gua gan etg gat gee gag ana ace gtg eeg gan man	30046

Asp Lys Thr Ile Glu Glu Leu Asp Ala Glu Lys Thr Val Leu Glu Lys 3035 3040 3045	
too awa gug gga gma cas tog ogo tit gat ago tat ego eka obg omt Ser Dys Ala Gly Ala Gln Ser Arg Phe Asp Ser Tyr Ser Dys Leu His 3050 3055 3060	30094
gat gam een afr eac goo ygD gwa aan nae got atg aog oto oga gog Aap Glu Aan 1le Aan Ala Gly Glu Aan Gin Ala Met Thr Leu Ary Ala 3065 3070 3080	30142
time gea god ggg off acc acg geg gft dag gom him; ogt off ged ggd Ser Ala Ala Gly Leu Thr Thr Ala Val Glin Ala Ser Arg Leu Ala Gly 3085 3090 3095	30190
gra ger get get ets gra eet ame ate tte gge tte gee ggt gyt ggt Ala Ala Asp Leu Val Pro Aso I.k. Pho Cly The Ala Cly Gly Gly 3100 3105 3110	30238
age out typ gap get ate get gag gon ace gag tat gts ang gas tit Ser Arg Trp Gly Als Ile Als Glu Als Thr Gly Tyr Val Met Glu Mie 3115 3120 3125	3 02 B6
tex: grd. eat git aig est enx: gam gog got oom oft age cam tot gam Ser Ala Ash Val Mat Ash Thr Glu Ala Asp Lys Ile Ser Glu Ser Glu 3130 3135 3140	30334
ace tac out one out can use tag use all one out sat sat god The Tyr Arg Arg Arg Chn Glu Tep Glu Fle Gln Arg Asm Asm Ala 3145 3150 3155 37.60	303 82
gea gog geg otg eaa cae old gel god dan ott eaa tog otg goe gla Glu Ala Clu Lou Lys Gln Leu Asp Ala Gln Leu Lys Ger Lou Ala Val 3165 3170 3175	30430
ego ont gam goo goo gta tig caa aaa acc ago oin ama acc caa caa Arg Arg Glu Ala Ala Val Jaw Oln Lys Thr Ser Lau Lys Thr Glo Cln 3180 3185 3190	3047H
gag cag suc CBS god cas tig god tir cig cas ogt 600 tid ago aat Glu Glu Thr Glu Ala Glu Leo Ala Phe Jan Glu Arg lys Phe Ser Aso 3195 3200 3205	30526
cas gog the two sac ten cha cet ego oga che exa eco att tac the Gln Ala Leo Tyr Asn Trp Leo Are Gly Are Leo Ala Ala Ile Tyr The 3210 3215 3220	30574
cam the two yax the get ate des est tet the alg see see est etc. Cln Phe Tyr Asp Leu Ala Ile Ala Arg Cyr, Leu Met Ala Glu Glu Ala 3225 3230 3235 3240	30622
ted cyt tgy gaa att. Ayn gat gad tot got ego tit all ama cog ggd Tyr Arg Trp Glu ile Ser Asp Amp Ser Ala Arg Ime ile Lys Pro Gly 3245 3250 3255	30670
ged tag can aga acc tat gos ggt. Atg Cly gos gat gas acc ttg alig Ala Trp Gla Gly Thr Tyr Ala Gly Leo Leo Ala Gly Glu Thr Leo Met 3260 3265 3270	אָרייתנּ
cta agt tig gea caa ally ywa geo goo cot tia aga cyn gwi ees ogo Lou Sor Loo Ala Cln Met Clu Asp Ala His Leu Ang Arg Asp Lys Arg 3275 3280 3285	30766
goa tha gag giv: gas cy: aca gia log olg god gas att tal got ggt Als leu Glu Val Glu Arg Thr Val Ser Leu Als Glu lle Tyr Ala Gly	30814

3290	3295	3300	
Lea Pro Glit Asp Sys	gge eea tte te Gly Pro Pho So 3310	x ctg acg caa gaa ab x Lou Thr Glu Glu 11a 3315	: 909 ୧୯୧୨ - 30 862 ድ Glu Ly s 3 32 0
olg glg ast gos ggt Leo Val Asm Ala Gly 3325	Ser Gly Ser Al	o: ygar agu ggt aat aa a Gly Ser Gly Asn As 3338	: aat ttg 30910 n Ason 3 enu 3335
ges tit gge ger gge Ala Hw: Cly Ala Gly 3340	acg gar art as Mu Asp Mu Ly 334	aa act bet lug ong ge Thr Ser Leo Glin Ali 15	a Sex Ile
toa tto got get tte Ser Leu Ala Asp Leu 3355	aam att ogt ga Lyn Jin Arg (1] 3360	ng gat tao ong gaa to nu Asp Tyr Pro Glu Se 3365	t all ggc 31006 r He Gly
ana ato ega egu ato Lya lle Arg Arg Ile 3370	: Dos cag atc ag : Dys Glm (),e Sc 3375	r gtt acc etg mng grj ir Val Thr Leu Pro Al: 3380	g ctattg 31054 g Leu Leb
Gly Pro Tyr Gln Asp 3365	Val Gln Ale I3 3390	a tta tot tac ggo ya Ge Lou Bor Tyr Gly As 3395	> Lws Āla 3400
gge tie gog aac gge Gly Leu Ala Asm Gly 3405	'Cys Ala Ala Le	ug god gdt tee eae gg au Ala Val Ser His Cl 3410	: ecy eat. 31150 y Thr Aso 3015
Asp Ser Gly Gln Pha 3420	:C in Le u Asp Pt 3 4 2		c Leu Pro)
lith gas yet aim goo Yee Glu Gly 11e Ala 3435	्र att gat caa gg . Ile Amp Glm सा 3440	pt acg cts aca ctg ag ly Thur Leu Thir Leu Se 3445	t tit cet 31246 r Phe Pres
Asn Ale Ser The Pro 3450	Ala Lye Gly Ly 3455	ao caa god act atg tu ys Gin Ala Thr Met be 3460	i Lye Thu
fen yan yab Ilo ile	. Ung cat att on Leu Nis I)e Ar 3470	nc tar acc att aag ta My Tyr Thr Ile Lys 3475	31336
coatecomae acagaact	AN BUCABBECCE (aercdddi criftaada	gtttet atq 31395 Met.
Gin Asu Ser Gin Thr	itto ago alg ao 'Pho Seo Met 11 3485	x: gag chg ten tha co nr Glu Leu Ser Leu Pr 3490	t aag gys: 31443 o Dys Gly 3495
ggo ggo got att acc Gly Gly Ala Ile Mu 3500	Gly Met Gly Gl	w goa tta acg eeg go lu Ala fau Thr Pro Al 3505	e gyg nog 31491 n Cly Pro 3510
est ngt aty qoa goo Agy Gly Myl. Ala Ala 3515	: 1.la tog etg to : 1eu Ser Leu Pr 352	raitte eec att tet go Foliasi Ero Ile Ser Al 20 - 352	a Cily Arg
99t fat gee eee teg Gly Tyr Ala Pro Ser 3530	rete ang ntg as Now Thr Lew As 3535	÷c tauc and age ggm ad sn Tyr: Asm Ser City Th 3540	2 ggl. aac: 31587 1 Gly Asn

age eeg tte ggt ete ggt tgg gae tgt aau gte atg aca att egt egt Ser Pro Phe Gly Len Gly Top Asp Cys Aso Val Met Thr Ile Arg Arg 3545 3550 3555	31635
oge ace and ace age gap one aat tat gal ace ace gat act tit etg Arg The See The Gly Val. Pro Asn Tyr Asp Ghi The Asp The Pho Leu 3560 3575	31683
gng orn goa got gan gtg ttg gtc gta gra tta mat gag gca ggt cas Gly Pro Glu Gly Glu Val læn Val Val Ala Leu Aen Glu Ala Gly Gln 3580 3585 3590	31731
get gat als: mys agt goa tee tea tis sag ggs als aat tig ggg alg Ala Amp Ile Amg Ser Glu Ser Ses Leu Clu Gly Ile Amu Leu Gly Met 3505 3600 3505	31779
and the ace git and gift tall ogo too out the gas age can the age. The The The Val The Gly Tyr Ang Ser Ang Lou Glu Ser His Phe Ser. 3610 3615 3620	31827
egg tig gas tan igg Cha our cha ana ana ggo gne aou gat tin igg Arg Leu Glu Tyr Trp Gln Pro Gln Thr Thr Gly Ala Thr Asp Phe Trp 3625 3630 3635	31475
ctg ata tac ago coo gar gga caa goo cat tta ctg ggr aan aat cot Leu Ile Tyr Ser Pao Asp Gly Gly Ala Nis Leu Lou Gly Lye Amy Pau 3640 3655	31923
can gon ogo all ago aat oos ota aal got ago caa aga gog caa tgg Gln Als Arg Ile Ser Aso Pro Leu Aso Val Aso Gln Thr Ala Gln Trp 3660 3665 3670	31971
eta tig gaa gur teg gia tea ten eac gge gag tag att tat hat eag Leu Leu Glu Ala Ser Val Ser Ser His Gly Glu Qla Ila Tyr Tyr Gln 3675 3680 3685	32019
tat oga god gas gad gas act gat tgd gas act gad gas etc aca gud Tyr Arg Ala Clu Asp Glu Thr Asp Cys Glu Thr Asp Glu Leu Thr Ala 3690 3695 3700	32067
cae ong ase are see gite cag one the etg caa gita gita cat the ggit Mis Pro Asm The The Val Glm Arg Tyr Lew Glm Val Val His Tyr Gly 3705 3710 3715	32115
And tota acc got age gas gta tit too acg cia ast ggm gni gai com Ann Leu Thr Als Ser Clu Val Phe Pro Tim Leu Asm Gly Ann Ann Pro 3720 3725 3730 3735	32163
ide aan Lot ggs tog ttg tte tig. the gta tit get tad gg. gag ogs Leu Dys Ser Gly Trp Leu Phe Cym Leu Val Phe App Tyr Gly Glu Arg 3740 3745 3750	32211
and and ago the tet gas alg dog one tit sas ged and agi, and tog lys Asm Sec Leu Ser Glu Met Pro Pro Phe lys Ala Thr Ser Asm Tro 3755 3760 3765	32259
clt tgc cgc aas gap ogt tit ten ogt lat gas tan ogt tit ges itg Leu Cys Arg Lys Asp Arg Him Ser Arg Tyr Glu Syr Gly Pim Ala Leu 3770 3775 3780	32307
ogo acc ogg ogo lin tot ogo caa ata otg ang lin ogo ogt otg caa	32355
Azg Thr Arg Arg Leu Cys Arg Glo Ile Lou Met Ehe Ris Arg Leo Glo 3785 3790 3795 ace etg tot ggt cag gco ama gg: gae gat goo eee ges tha glo tea	

Thir Len Ser Gly Gli Ala Lya. Gly Asp Asp Glu Pro Ala Len Val Ser 3800 3805 3810 3815	
ogt utg a(a ctg gat tat gac gan and gog gtg gtd agt aog etc gtt Arg Leu He Leu Asp Tyr Asp Glu Asp Ala Val Val Ser Thr Leu Val 3820 3830	32451
tet gte uym ega gfg gga eat gan eaa gat gge aca aeg geg gte gee Ser Val Arg Arg Val Cly Ris Clu Gla Asp Cly Thr Thr Ala Val Ala 3835 3840 3845	32 4 99
ctg ccg cca ttg gaa ctg get tat cag cet ttt gan cca gaa caa aaa Leu Pro Pro Leu Glu Leu Ala Tyr Glu Pro She Glu Pro Glu Glu Lys 3850 3855 3860	32547
gea etc tgg oga eta atg gal gla etg gog aat the aac acc atc caa Ala Leu Trp Arg Pro Met Asp Val Leu Ala Asn Phe Asn Thr Ile Ghn 3865 3870 3875	32595
ege tag cas etg ett gat etg cas gge gas gge gts eee ggt att etg Ang Top Glo Leu Leu Aep Leu Glo Gly Glu Gly Val Pro Gly 11e 1eu 3885 3890 3895	
tat cag gen ass ast ggm tog tog tat ogs tot get cas egt cag ace Tyr Glo Asp Lys Asm Gly Trp Trp Tyr Arg Ser Ala Glo Arg Glo Thr 3900 3905 3910	3 2 691
GOU See You aby eat gog gite ace tog goe ame atg camete cit cet G)y Glu Glu Met Amu Ale Vel Thr Trp Gly Lys Mes Gin Lou Leu Pro 3915 3920 3925	
ato any occ got att ong gat aan got toa otg atg gat att est ggt. Ile Thr Pro Ala Ile Gln Asp Asn Ala Ser Leu Met Asp Ile Asn Gly 3930 3935 3940	
DAT 999 can cluy gat ten get atc acc get con gon cta and got tat Asp Gly Clu Leu Asp Trp Val Ile Thr Gly Pro Gly Leu Arg Gly Tyr 3945 3950 3955	
can ago cag cat oca get ggo agt tgg aca ogt till ang tog titg oco His Ser Gln His Pro Amp Gly Ser Trp Thr Arq Phe Thr Pro Leu His 3960 3965 3970 3975	32883
gur the con sta gas hat som cat our one goe can out gon gat the Ala Leu Pro lle Glu Tyr Thr His Pro Arg Ala Gln Leu Ala Asp Lou 3980 3985 3990	32931
ang ggg geo ggg eng hrx: gal tha ghe eng aft ggb eoc ama ago gng Met Gly Ala Gly Lau Ser Asp Leu Vol Leu Ile Gly Pro Lys Ser Vol 3995 4000 4005	329 79
ogt tig tat gee ast asc ogt gat ggr. til. acc gea ogs ogg gat gig Arg len Tyr Als Asn Asn Arg Asp Gly Phe Thr Glu Gly Arg Asp Val 4010 4015 4020	33027
guy cas the ggf. ggf. and each old ong the ong mme nee gat geg ogt Val Glm Ser Gly Gly Ile Thr Leu Pro Leu Pro Gly Ala Asp Ala Ang 4025 4030 4035	33075
aag tta gtg god til age gae gta ete ggt tea gge eaa gea mai tilg Lys Leu Val Ala Phe Ser Asp Val Leu Cly Ser Gly Cln Ala His Leu 4040 4045 4050 4055	33123
gtt gaa git agt geg aeg aas gid see tee tee tee eas eat eig ges eat Val Glo Val Ser Als Thr Lys Val Thr Cys Trp Pro Asn Leu Gly His	33171

4060	4065	4070
egg) ogt til ggt cog com	ato are the mny ega tit	age caa tee gee 33219
Gly Arg fhe Gly Gln Pro	The The Leu Pro Gly the	Ser Cln Ser Alø
4075	9080	408S
god Ael thi aat ooi gat	ona git cat obg go: gal.	rlg gar ggt agt. 33267
Ala Aen Mom Aen Puo Nep	Ary Val Die Leu Ala Asp	leu Asp Gly Ser
4090	4095	alon
Gly Pro Ala Asp Leu Ile	tat gil cel gol yez cec Tyr Val Rie Ale Asp Bis 411.0 4115	ety get att tte 33315 Leu Asp Ile Phe
age aat gae egt ggw eac	ggt tit ges cas ces ite	aca ete egt tit - 33363
Ser Aso Glo Ser Gly Aen	Gly Phe Ala Gin Pro Phe	The Leo Arg Phe
4120 4125	4130	- 4135
oct gao ggo cty ogh tit.	get gat act tyc cag cta	caa gtg gct gat 33 4 11
Pro Amp Gly Leu Arg Phe	Asp Asp Thr Cys Gin Leu	Gln Vol Ala Asp
4740	4145	4150
	gto ago otg ato otg ago Val Sor Law Thu Teph Ser 4160	
guy cræ car cat tgg egc	tge gat etg ace aac geg	ame con tyn tta 33507
Ala Pro Ris Hic Trp Any	Cys Asp Leo Thr Asn Ala	Lys Pro Trp Leu
4170	4175	1180
Leu Sor Glu Met Asa Asa	aac atg goo got cat cac Asm Met Gly Ala His His 4190 4195	ace etg cat tac 33555 Thr Leu His Tyr
ogt age too ghe dag tht	tgg cóg get gee een gox;	gon gno baa got 33603
Ang Ser Ser Val Gla Phe	TTP Leu Asp Glu Lys Ala	Ala Ala Leu Ala
4200 4205	4210	4215
acc ggs cas ace org gto	tgt tae etg eer tte eeg	gto cat acc ctg 33691
The Cly Gin The Pro Wel	Cys Tyr Leu Pro Phe Pro	Val Bis Thr Len
4220	4225	4230
tigg coar acu gair acc gag	gat gam ate age gge amt	ann tta gtg acc 33699
Trp Cla Thr Glu Tir Glu	Anp Clu Ile Ser Gdy Arm	Jys: [49] Vol Thr
4235	4240	4245
act tha ogt tee get tee	And got tog got man ogt	gag egg gaa tit 33747
Thr Leu Ang Tyr Ala His	Gly Ala Top Amp Gly Arg	Glu Ang Glu Phe
4250	4255	1260
Arm Gly Fire Gly Tyr Val	ලකු නැද කෙන ඉතිර කුල eat Glu නියා නිහැ එකුව පිනැ His 1270	can etg get com 33795 Glin Leu Ala Glin
gge aat geg ceg goa egt	aca too bog goa oft acc	ama aac tgg tat 33843
Gly Asm Ala Pro Glu Arm	The Sur Iro Ale Len The	Lys Amn Typ Tyr
4280 4285	4290	4295
goe ace gga ate cet gag	gta gac aat acy cta tet	ged ggg tat tgg 33891
Ala Thr Cly The Pro Clu	Vel Asp Asm The Low Ser	Ala Gly Tyr Typ
4300	4305	4310
ege ggt gat aog cag get	the act ggt thi acg era	can tit act etc 33939
Arg Gly Asp Thr Gln Ala	The Thr Gly Phe Thr Pro	Nie Phe Thr Leu
4305	4320	4325

Tro lys Glu Gly Lys: Asp Val Pro Leu Thr Pro Glu Asp Asp Ris Ass 4330 4335 4340	
cty (an igg tis aac tog gou oto amp gg) caa oca etg egi agt gas Leu lyr Trp Lou Ann Ang Ala Leu lys Gly Gln Pro Leu Ang Ser Git 4345 4350 4355	a 34035 i
Cic tac ggg cta gat ggc ago gus rag cag aag atc ccc tat aca gtg Leu Tyr Oly Leu Asp Gly Ser Ala Gla Gla Lys Ito Pro Tyr Thr Ve 4360 4365 4370	
act gas too my: one caa gtg ogd caa tta das gat ase eet med ett Thr Glu Ser Ang Pro Glo Val Ang Glo Leu Glo Asp Aso Thr Thr Lei 4380 4385 4390	
tou the gird off the gen too gird gow agt out agt thit can bar Ser Pro Val Lou Tru Alm Ser Val Val Glu Ser Ard Ser Tyr His Tyr 4395 4400 4405	r
988 Off atc atc ago gat occ cas the east cap gat atc act obly to Glu Arg Ilo Ilo Sex Arap Pro Glo Cys Aen Glo Amp Ilo Thr Leu Ser 4410 4415 4420	r
agt gac ctd the gag cas one of sas cay gift the gift cas this on Sar Asp Leu Phe Gly Glin Pro Leu Lye Glin Vol Ser Vol Glin Tyr Pri 4425 4430 4435	Þ
ogr oge aat aan een nen aet eng tat ook gal aca ela een gal Arg Arg Arm lys Pro The The Arm Pro Tyr Pro Arp The Leu Pro Arg 4450 4450 4450	> 5
act our tit goo ago agt tat gao gao cas cas cas cis thy cag the The Leu Phe Ala Ser Ser Tyr Amp Amp Clin Clin Cau Lau Amp Lei 4460 4465 4470	п
and the day can bee agt toy cat cat cha att get aut gun etc ag The Tyr Cln Cln Set Set Trp Ris His Len He Ala Aen Cln Len Ar 4475 4480 9485	ğ
gfy tha ggs tha dog gat ggt ada dgd agt gat ggt the act the ggs Val Lou Gly Lou Pro Asp Gly Thr Ang Ser Aup Ale Phe Thm Tym Asp 4490 4495 4500	P
get aan eac gig oet git gat gijt tin aat eig gra get ein igt ge Als bys His Val Pro Val Amp Gly Leu Ash Leu Glu Ala Leu Cys Al 4505 4520 451.5	a.
gas sall age ong all god gat gat ass out ogd gas tad one dad da Glu Amn Ser Leu Ile Ala Asp Amp Lys Ero Amn Glu Tyr Leu Ash Gl 4520 4530 453	n 5
CBS Cya acq ite tat acc gai gyg aaa acc gai gyg ann ann acc do do Glin Arg Thr Pho Tyr Thr Asp Gly Lys Thr Asp Gly Lys Asp Pro Th 4540 4545 4550	r
ora ctg amm aca ccg aca cga cag gct tta mic qcc tit acc gan ac Pro Leu lws Thr Pro Thr Arg Cln Ala Leo He Ala Phe Thr Glu Th 4565 4560 4565	r.'
gry gia tia nog gas tot ong tia too gos hit get gg: ygl sin so A)e Val Leu Thr Glu Ser Leu Leu Ser Ala Phe Asp Gly Gly 11e Th 4570 4580	g 34707 r

can get ges tis con gga ott utg een can ges gga isn cen cen cen gen Pro Asp Glu Leu Pro Gly Leu Leu Thr Glu Ale Gly lyr Glu Glu Glu 4585 4590 4595	34755
cot, tak ong tito ona ono agit gap gaa aan daa gib ing gita goo ogi Pro Tyr Leo Phe Pro Leo Ser Cly Glu Aso Glo Val Top Val Ala Arg 4600 4605 4610 4615	34603
aga ggo tak son gat tao gga act gag gta (sa tit tgg ogt oct gio Lys Gly Tyr Thr Asp Tyr Gly Thr Glu Val Gln Fhe Trp Arg Pro Val 4620 4625 4630	34851
gen can egt ame ace eag til a see gyg ama acg act etn ann bgy gat Ala Glu Arg Aso The Glu Leu Thr Gly Lys The The Leu Lys Tep Asp 4635 4640 4645	34899
ace cae two type gic ate act (00% Acc come gae geg get gg) big act The His Tym Cys Val lie The Glo The Glo Asp Alm Alm Gly Leo The 4650 4655 4660	34947
gto toa ger aat tat gad tog ogt tit oto aca oot aig om oig act Val Sor Ale Ash Tyr Asp Tip Ang Phe Len Thr Pro Net Glin Leu Thr 4665 4670 4675	34995
gat atu ken gen kut gitg dai aid aia ann tig gai gog dia gga ogd Asp lle Asm Asp Asm Val His lle lle Thr Leu Asp Ala (eu Gly Arg 4680 4685 4690 4695	35043
oct gite act caa egi the teg gga ate gaa ast ggi gig gea acs ggi. Pro Val The Gin Arg Phe Trp Gly Tie Glu Asm Gly Val Ala The Gly 4700 4705 4710	35091
two tot top one goa gos eas now the act one one gho get give earl Tyr Ser Ser Pro Glu Ala Lys Pro Ehe Thu; Pro Ino Val Asp Val Ash 4715 4720 4725	35139
got god wit you duy acc ogs com our cot gto gog cag tgi dig gto Ala Ala Ile Ala Leu Thr Gly Pro Leu Pro Val Ala Gla Cys Leu Val 4730 4735 4740	35187
law gog cng gax; agt tog atg ong otta the ggt ong gen acc the amo Tyr Ala Pro Asp Ser Try Met Pro Lew Phe Gly Olm Glu Thr Phe Am 4745 4750 4755	35235
aca tha acy cop gas gag cas sag aco cho cgt gat the cop ath atc Thr Leu Thr Gln Glu Glu Gln Lys Thr Leu Arg Asp Leu Arg Ile Ile 4760 4765 4770 4775	35283
ace gas get tog oot att tog ook oby oot one ook ook ook ook ook ook ook ook ook oo	35331
agt cas ass got age aca oca the gtt dag etg tta ach aso ago ate Ser Gln Lym Ala Gly Thr Pro Leo Val Lym Leo Leo Thr Ash Ser Tic 4795 4800 4805	35379
ggt tta oot ooc cae as; oto atg otg got ang gan ogt tat gac ogt Gly Leu Pro Pro His Amn Leu Mel. Lou Ala Thr Amp Ang Tyr Asp Ang 4810 4815 4820	35427
gat tot gaz cag caa att ogt caa caa gto gos Ulu agt gat ggt tit Asp Ser Ghu Gho Gho The Arg Gho Gho Val Ala Phe Ser Asp Gly Phe 4825 4830 4835	35475
THE OUT IN THE CAN GOT HET BY ONE CAL HOW HER GUT HAN BOY THE	35523

Ciy Arg Leo Lou Gin Alo Ala Val Arg His Giu Ala Giy Giu Ala Trp 4840 4845 4850 4855	
can out asc can get got tot otg gtg aco ann atg gon got acc ann Glin Arg Ann Glin Asp Gly Ser Leu Val Thir Lys Met Glo Asp Thir Lys 4860 4865 4870	35571
seg ege tgg geg att aeg gga ege aet gaa tat goe aat aag ggg eag Thr Arg Tro Als Ile Thr Gly Arg Thr Glo Tyr Asp Asm lys Gly Gla 4875 4880 4885	35619
gog eta oga adl fak dag dom hat blu obd aak gad ligg oga tat gog Ala Ile Arg The Tyr Glo Pro Tyr Phe leu Asm Asp Tro Arg Tyr Val 4890 4895 4900	35867
agt gat 990 agn ags ees gag grow tot goo gat ont cat ato tot Ser Amp Amp Ser Ala Ang lys Glu Ala Tyr Ala Amp Thr His Ile Tyr 4905 4910 4915	35715
gat cog att ggg Cgg god atc caa gtt atc acg gca aaa ggc tgg ctg Asp Pro Ile Gly Arg Glu Ile Glu Val Ile Thr Ala Lys Gly Trp Leu 4920 4935 4930 4935	35763
ogg cag aar caa tat tto oog igg tit soo gtg agt gaa gat gaa aat Arg Glu Asn Glu Tyr Hoa Pro Trp Pha Thr Val Ser Glu Asp Glu Asn 4940 4945 4950	35811
gat tig too got gae gog ete gig taa tigaateaag attogetegi. Asp Leu Ser Ale Asp Ala Leu Val 4955 4960	35858
ttaatgttaa ogagogaata taatateeet aatagattte gagttgeege goggoggeaa	35918
gtgaacgaat coccaggogo atagataact atgtgactgg ggtgagtgaa agcagccaan	35 9 76
asagoaycag chigawagat gaagggtata aataagaaso tgcattgtga gttota <u>aata</u>	36038
gagtagoago atatottati gootelloati, koslaggiaa taassitoss tigotgiaaa	36098
aatotgtost catgagaact aaaastaaca addittoimil dignaagaga satosataat	36158
tozattaaaz atgitataga atrigaaloa agacmatthy thegotoato vaaamiatam	36218
acateogost egytaataas agetgatgin astayaastil lettititat oocaagigae	36278
atatyteest setematace agastastta gatataceaa aaccatilau etagtaatet	36338 '
aattgatatt ttaaattact ttteetataa egetgaetta aattaakeen auurattood	36398
gtgatgaaat tataaaagtt aacattatoo gatagataas aaccatgotg Edgeasacta	36458
aatoggetet titeteeett tittataasa tiaaccette etillilikuke etiatitaea	36518
cragcaatan tigasagass agregotits coogcordat taxetooge datacogniti	36578
Helixiaacmi gaasakwim attgattott gaaaaantag tottaccatt aataacaacc	36638
inisacras Laactikaay cateestase eestassat saegtassaa agaasatase	36698
novaecasta ataatiticg tylalinggt totoanoaga gasatagaag asecsataat	36758
agaaagaaaa gootaagata aaaatataat ceceggeeen gestoncom coognaagoo	36818
ASSASSIASON DESCRIPCIO ASSASSASSAS CASSASSAS CONTRACTOR ASSASSAS E	7.5005

tttoograca gataammagi loggacaamta tqammagataa tttatttoom tatatgatag 36938 attatament amcamcatgo atalenatam amcamcactg gentatatta atgatatata 36998 atcogocilig titigggatti grigawwygy: artiticacat aataqatata aaaqcagaac 37058 agotakhynn ataataagga negamathik attittatta aaacaataac gaagattoat 37118 tataleeggo aatgazmaza emudigatga saataattit ttattiotat taattatata 37178 exeluyytyty abattuese, efestatosa tyotactaat pyantaacta atytaasaat 37238 caaatostat kalinthooso tootgaatga tgoogoonga agasagaska usquanusal 37298 esasenatgo assessotte ettomastan pominatore attacegose acquatactet. 37358 сававляли: асадатдава датальности отлатьяльный thicongless авиалостат 37418 aangaayaka ataactatog gamangozot atxaatekan kasangalan gantaasaan 37478 саноуtitti itaootaoon андлинеуы. yetiyestic teetiiyosy aaggaaassa 37538 ecttatytta atempuksaa akaematata taemattama yatatyymay taasataasa 37598 tgattitutg (sgccstcig gastaataat attggaagat aaagttatta aaacctcaaa 37658 gataccacty wouldbycop yasytestas aagaaaasyg satataatgs cattititati 37718 eccagacyca cattlettta testaeetti atattecaag gesteagoga tiattaaatt 37778 retartgest stotamests amentolese teatgicstit ggigeatoit tegggeatiit 37838 ogtootgyma tyranakata aatsyttaot gaaaacaata cattgattit taattasata 37898 37948 otygogálat gacottaatg atgotacttt attitooxyt attoxating

<210> 12

<211> 954

212> FRT

<213> Photorhabdus luminescens

<400> 12

Met Lym Ash lle Amp Pro Dys Leu Tyr Gln Lys Thr Pro Val Val Amn 1 10 1.5

The Tyr Aug New Yang Gly Lem Thu Ile Arg Asm Ile Aep Phe Uis Arg 20~~25~~30~

The The Ala Ash Gly Asp The Asp IIc Arg IIc The Arg Ris Clm Tyr 35 40 45

Asp Ser Lea City His Lea Ser Gin Sor Thr Aup Pro Arg Law Tyr Gia 50 60

Also Lys C(n Lys: Ser Arm Piet Leu Trp Gln Tyr Asp Leu Thr Gly Asm 65 70 75 80

tie Leu Cya Thr Glu Sor Vol Asp Ala Ciy Ang Tio Val Thr Leu Arab 95 90 95

ASD The Chu Gly Arm Pro Leo Leo Thr Val Thr Ala Thr Cly Val The 100 105 110

Gin Thr Arg Gln Tyr Glu Thr Ser Ser Leu Pro Gly Arg Leu Leu Ser

		115					120					125			
A9J	Thr 130	Glu	Gln	Ile	Pro	Glu 135	Ľγs	יולני	Ser	Arg	11e 140	Thr	Glu	Arış	Lou
11¢ 145	נלטינו	λla	Gly	A ≘n	Ser 150	Glu	Ala	Glu	[ge;	A(a) 155	нія	Asn	TÆI	λla	ളക്ക 160
Gln	Cys	Val	Arg	н <u>і</u> я 165	туг	asp	Tìm	Ala	Gly 170	ИЭЛ	JJJ.	prá	Leu	Glu 175	Ser
Leu	Ser	Leu	Thr 180	Cily	Tio	Val	Leu	Ser 185	Gln	Ser	Ser	Gln	Leu 190	Leu	Se1
Asp	'line	G) n 195	ផាប	Νlė	Ser	Ттр	Tha: 200	Gly	Αæp	Aem	Głu	'Ihr 205	Val	Ттр	Gln
Agn	16et 210	स्या	Ala	A s p	ASO	11e 215	Tyr	Τπ	Thir	تجها	8en 220	Ala	Phe	Mego	AlA
Thr 225	Cl y	Ale	Leu	Leu	Tîm 230	ദ്ധമ	'lhr	Автр	Ala	18/5 235	Gly	Asn	Ile	Gln	Arg 240
t.eu	Thr	Tyr	Лер	Val 245	УĴЪ	Gly	Glu	Leu	Аво 250	Gly	Ser	Ттр	Leu	The 255	Leu
Lyz	Perp	Gln	Pro 260	Glu	Cln	Val	Ле	T.J.© 265	gnA	Sear	(<i>e</i> s)	Thr	Туг 270	Ser	Ala
AJa	Gly	Gln 275	-	Leu	YLÜ	Glu	Glu 280	His	Gly	Æn	Gly	Val 285	ile	Thr	Glu
<u>ም</u> ያድ	Ser 290	_	Glu	Pro	GJu	Thr 295		gln	ren	Ile	300 <i>GJ</i> À	Thu	Lys	The	His
Ах 9 305		Ser Ser	увар	уJэ	1ye 310		Leu	Glin	Asp	16€0 315	_	туr	Glu	Тут	As p 320
Pro	Val	Gly	Pist)	Va1 325		Scr	Πe	Arg	Ажт 330		Ala	GLu	Ala	Tha: 335	Ахц
Phe	drt.	Hie	340		lys	VS)	Ala	Pro 345) Aen	Time	Tyx	Thi 350	_	лер
Ser	Leu	Тут 359		Lou	ılle	Ser	A3a 360		· GLy	Arg	Glu	M et 365		Asn	Ile '
Gly	Gln 370) Ser	AET.	Gln	1 <i>e</i> 0 375		Sen	Leu	TÎTA	Leu 3HD		Sen	Asp	ne.A
Aso 389		· Tyr	The	lea.	390 1771		· Arg	Tene	Туг	. Ting 395		Asig) Arg	Ghy	61y 460
Next	ie:	Thu	. Lye	11e 405		llís	: Ser	Ser	Pro 410		. Tilor	. с уп:	AEI:	415	тут
Thu	Thu	. Asc	420		· Val	. Ser	AET.	425		: Asr	ALC:	Alö	Val 430		Sex
'Ita	Tes	1 Tro 43:		i Neg	e Pro	ALE	G(r 440		Þщ	ι Αξε	i IÆ	Pive 445	_	Ala	GLy
Gl	/ Hi: 45(, ya	13 10	i Leo	11e 459		Gly	/ Gla	n Asar	1 Lou 460		TI	a Afir	Than

Arg 46 5	Oly	Clo	[au	Cln	His 470	Val	l'nr	Leu	Va)	Lye 475	Arg	Aep	Ly∞.	Сlу	Ala 480	
Aen	Asp	Asp	Arg	Glu 485	Тира	Тут	Arg	Тут	Scr 490	Ser	ASO	GΙУ	Arg	Arg 495	The	
Tear	Lys	Ilo	Asn 500	Glu	Gln	erv	Thu	Ser 505	Ser	Arm	िदर	Cln	Thr 510	ദ്ധ	ул.д	
ιŀe	Thr	1yr 515	Leu	Pro	Ser	Leu	Glu 5 20] trai	ÀΣΤΙ	L∉u	"ከድ	Gla 525	Asn	Ser	Than	
IJē	Thr 530	'l'me	Glu	Asp	Leu	Gln 535	VaJ	tae	Thr	Val	G1γ 540	Glu	Ala	ЗΥ	Arg	
Ala 545	G h ri	Val	Arg	Vət	Løn 550	Ris	Түр	AEP	Ser	Gly 555	Gla	Pro	Glu	Asp	11e 560	
ASD	Asn	Asn	Gln	ьен 565	Arg	T ንጥ	Ser	Туг	Asp 570	æn	Leu	Ile	Gly	9ec 575	Ser	
Gln	Leu	Glu	Lец 580	Asp	Sor	Lys	Gly	Glu 585	Ile	IJe	Ser	ØLu	Glu 590	Glu	тух	
Tyr	Pro	Тул: 59 5	Giy	Cly	בולנו	Ala	Leu 600	Trp	Мą	The	Arg	605 605	Arq	Thr	Glu	
Ala	Ser 610		ъув	Thr	IÌ¢	Ang 615	ገን፫	Ser	Gly	Lye	Głu 620	Arg	Asp	Ala	'Itar	
Gly 625	نها	Tyr	Tyn:	ፕ ሃェ	Gly 630	Tyr	Arg	T ሃ <u>Y</u>	ብ ን.	Gln 635	Pro	'l'rp	Val	GJA	Arg 640	
Тър	والجما	Ser	Ala	Авр 6 4 5		Ale	Gly	Tin	Val 650	Asp	GJX	Leu	Asan	L ena 6 55	Tyr	
λrg	Met	Val	Arg 660		A eti	Pru	Val	Thr 665	Leu	Leu	лер	Pzo	А эр 670		Leu	
Met	Pro	7far 675		λla	Glu	Arg	lle 6KD	Ala	Ala	Lag	(In	Lys 665	Asn	lys	Vel	
Mle	Асд) 690		Ala	ľTO	Ser	P ra 695	MJX.	Asn	Ala	The	700		Ala	Πe	Asn	
705	Arg	Pro	Pro	Val	እ <u>ነ</u> ኤ 710		Σув	Pro	Tine	1es 715		Lys	Ala	Ser	Thor 720	
Ser	Sec	Эjn	Sec. (Thu: 725		J,ĀZ	Pro	Ile	1 ₈ 93 730	Ser	Ala	Ser	Ile	1 ₈ /8 735	Pro	
Thr	The	Ser	Gly 740		වර්ග	Il∈	Tru	Ala 745		l <i>i</i> na	Sur	lro	Val 750	_	श जा	
Lys	Sor	Tta: 755		Glu	IJe	Ser	Ten 190		Glu	Sær'	Tra	765 765		Aso	Ser	
Ser	9em 770		Türg	Ser	'Ifn'	λ <i>е</i> п 775		हाम	lys	Lys	2e1 780		Ðæ	Lou	JAr .	
Arg 785		Αυτρ	Àsn	Arŋ	Ser 790		Glu	Asp	Met	ملا5 795		lys	Phe	ĘTO	HQD HQD	
Gly	Phe	Lys	Ala	Trp 805		Pro	Leu	. Asp	77a 810	Lycs	Net	Ala	Arg	Gln 815		

Ala Ser Val Phe Ile Gly Gln Lya Asp Thr Ser Aso ໂວນ Pro Lys Glu 820 825 830

Thu Val Lym Asn The Asn Thu Tup Gây Thu Lym Pro 1993 (ee Asn Asp 835 840 845

Low Ser Thr Tyr Ile Lye Tyr Thr Lye Asp Lye Ser Thr Val Trp Val 850 $\,$ 860 $\,$

Ser Tim Ala Tio Asi Tim Glu Ala Gly Gly Gln Ser Ser Gly Ala Pro 865 870 875 880

Leu His Glu Ils Asn Met Asp Leu Tyr Glu She Thr Ils Asp Gly Gln 885 890 895

lys Len han Pro Leu Pro Arg Gly Arg Ser Lys Asp Arg Val Pro Ser 900 905 910

led led for Amp Thr Pro Clu Ile Glu Thr Ala Ser Ile 1le Ala Led 915 920 925

Ash His Gly Pro Val Ash Asp Ala Glu Val Sor Phe Leu Thr Thr 1,273 930 940

Pro Leu Lys Asn Val Lys Pro Tyr: Lys Arg 945 950

<210× 13

<211> 2522

<212> FRT

<213> Photorhabdus luminosogra

<**400**~ 13

Wet The Leu Lys Gly The Ash Met Ash Ser Pro Val Lys Glu The Pro 1 5 10 25

Asp Val Lau Lys Ile Gln Cys Gly Phe Gln Cys Leu Thr Asp Ile Ser 20 25 30

His Ser Ser Phe Asn Olu Phe His Gln Gln Val Ser Glu His Leu Ser 35 40

Trp Ser Glu Ala His Asp Leu Tyr Bis Asp Ala Gln Gln Ala Gln Lys 50 55

Asp Amn Arg Leu Tyr Glu Ala Arg The Leu Lys Arg The Asm Pro Gln 65 70 75 RD

Leu Gin Astr Ala Val His Leu Ala Ile Val Ala Pro Am Ala Glu Leu 85 90 95

Ilo Gly Tyr Aso Aso Glo the Ser Gly Arg Ala Ser Glo Tyr Val Ala 100 105 110

Pro Gly Thr Val Ser Ser Mot the Ser Pro Als Ala Tyr Lou Thr Glu 115 120 125

kmu Tyr Arg Gio Ala Arg Asm Leo Hie Ala Ser Asp Ser Val Tyr Arg 130 135 140

len Asp Thr Arg Arg tro Asp Len Lys Ser Mot Ala Len Ser Gln Gln 145 150 155 156 Aen Met Asp Thr Glu Leu Ser Thr Leu Ser Leu Ser Asn Glu Leu Leu 2.20 len Glo Ser Ile Lys The Glo Ser Lys Leo Asp Asm Tyr Thr Glo Val **1**85 met Glu Met Law Ser Ala Phe Arg Pro Ser Gly Ala Thr Pro Tyr His Asp Ala Tyr Glu Asm Val Arg Lys Val Ile Gln Les Cln Asp Pro Gly leo Gio Gin Leo Asn Ala Ser Pro Ala Tie Ala Gly Leo Met His Gin Ala Ser Leu Leu Gly The Asm Ala Ser The Ser Pro Clu Lou The Asm He law Time Gio Gio Tie The Cio City Aso Ala Gio Gio Leu Tyr Lys Dys Asm Inc Cly Asm lle Glu Pro Ala Ser Leu Ala Mei. Pro Glu Tyr Leu Árg Arg Tyr Tyr Asn Leu Ser Asp Glu Glu Leu Sur Gln Phe 13e 295 Gly Lye Ala Ser Asn Phy Gly Gln Gln Glu Tyr Ser Asn Asn Oln Lea 315 Ilo Thr Pro Ile Val Asm Ser Asm Asp Gly Thr Val Lys Val Tyr Arg the fire Arm Giu Tyr Thr Thr Ash Ala Ash Gin Val Amp Val Glu Les Phe Pro Tyr Cly Gly Glu Asin Tyr Clin Lou Asin Tyr Lys Phe Lys Asin 365 360 Ser Arg Glin Asp Val Sir Tyr Leu Ser Ile Lys Leu Asn Asp lys Arg Clu low He Arm the Glu Gly Ala Pro Cln Val Asa the Glu Tyr Ser Glu Ris Ile Thr Leo Sor Thr Thr Asp Ile Ser Glo Pro Pho Glu Ile Gly Leu Tur Arg Val Tyr Pro Ser Ser Ser Top Ala Tyr Ala Ala Ala 425 Lye Fhe Thr Ile Glu Glu Tyr Agn Gln Tyr Sex Phe Lou Leu Lye Leu Asm Lys Ala lle Arg Leo Sor Arg Ala Thr Glu Leo Ser Pro Thr lle Tem Clu Ser Ile Val Arg Ser Val Aon Cln Gln Leu Asp lle Asm Ala

Clu Val Leu Gly Lys Val the Leu Thr LyB Tyr Tyr Met. Glo Arg Tyr

Ala Jie Asn Asa Chu Thr Ala Leo Ile Leo Cys Asn Ala Leo Ile Ser 500 505 510

Gìn	Arg	<i>S</i> er 515	Тух	Asp	Asn	Gln	P170 520	Ser	Cln	Pho	Аэр	Arg 525	Leu	i,ps	Asn
Thr	ያτο 5 30	Leu	Leni	Asar	θlλ	G)n 535	Тух	Plue	Ser	Thu:	GLy 540	Asp	Gl u	Glu	$\mathbf{Il}\mathbf{e}$
Asp 545	Leu	REI I	PYO	Gly	<i>S</i> er 550	Thr	Gly	Asp	mp	Arg 555	Lys	Ser	Val	Leu	Lys 560
λrg	ИS	되노	ABD	11e 565	Аер	Asp	Ile	Ser	Leu 570	Тух	λrg	د وخير آ	Tæn	Lys 575	De
Tht:	A (a)	Kis	Asn 580	ስ <i>ድ</i> ክ	Gln	λεφν	Gły	1.yg SB5	Tie	Lys	Asan	Asn	Leu 590	Asn	AET)
Leu	Ser	Аар 595	Leu	Тут	lle	Gìy	Lys 600	Leu	Leu	Ala	Glu	11e 605	His	(3)m	Lou
Tim	610	yab	Glu	Leu	Аер	1en 615	Leu	Lean	Val	Ala	Val 620	Gly	Glu	Gly	Glu
Tha 625	Дер	الجا	Ser	νje	X1¢ 630	Ser	Asp	lys	Gln	Leu 63S	Ala	УĴЭ	Leu	Ile	Л гэ 640
Lyb	Leu	Asn	Thar	Ile 645	'Itar	Val	Trp	Leu	Gln 650	Time	Gln	Lys	Tip	Scr 655	Alzı
Ehre	Gli	Leu	<i>P</i> he 660	Vel	Met.	The	Ser	Thr 665	Secr	Тухг	Asn	tys	Thr 670	Leu	Thr
PTO	Glu	11e 675	lys	A e n	Leu	Leu	Д ар 680	Thr	Val	Тут	HLS	Gly 685	رضا	വം	Gly
Pbe	Asp 690	БУЗ	Asp	ΓΆS	Ala	Asn 695	leu	Leu	His	Va)	Met 700	Ala	Pro	Эyr	Ile
Ala 705	Ala	Thr	Leu	GJI3	Leu 710	Ser	Ser	GJ.u	Astr.	Val 715	Ala	His	Ser	Val	Leu 720
Leu	Ttp	A)a	Assp	โพช 7 25	leu	Dys	PYD	Gly	Азр 730	Gly	Ala	Kieft.	Thr	Ala 735	Glu
Lye	Phe	Trp	Аз р 740	Trp	Leni	Asara	Tu-	Gln 745	Tyx	Thr	Pro	λер	Sex 750	Ser	Glu
Val	Īæi	Ale 755	TÎLY	Ğln	Clu	His	11e 760	٧al	Gln	Тут	Сув	Gln 765	уја	leu	Дla
Gln	Leu 770	Gλυ	Mat	Val	Тут	Нiв 775	Ser	Thar	Gly	ΣJė	Asar 780	Clu	Asen	Ala)†he
Arg 785	Lou	Ptac	Val	Thr	Lys 790	Pro	Glu	Met	Phe	ങു 795	Ser	Ser	ጥዢ	Glu	Al.a 800
V±1	Pro	Мa	nis	Ֆ որ 805	Ala	Leu	Ser	Leu	Ile 810	Met	Leu	Thu	Лrg	Phie 815	Ala
Авр	ТТЪ	Val	ക്ടന 820	Ala	Leu	Сîу	Glu	Lys: 825	Ale	Ser	Sect	Val	L σ\(830	Ala	Ala
Phe	Glu	Ala 835	Aan	Ser	Leu	Tim	ծ 840	C1u	C) TI	lktu	Ala	Asp 845	Ala	Met	Aem

ino Asp Ala Asm Leo Leo Leo Glm Ala Ser Thr Glm Ala Glm Asm His

	850					855					860				
Gln 865	His] eri	Pro	Pro	Va 1 870	Ħìr	Gln	Lys	Æනා	ala 875	Phe	වසර	Cyrs	מתג	Utr HRO
Sen	Ile	λер	7fm:	11e 885	Leu	CJV	TIP	Val	A≘n 890	Val	Ala	сľи	Gln	Leu 895	leja i
Val	Ala	Pro	ලා 900	Gly	Val	Seen	Мa	Len 905	Vel	ឲ្យវិទ្ធ	ſĘ'nĮ	ሊ ም	Tyr 910	Ile	Gln
Γ⊜π	A.an	3) n 915	Lyes	Jle±	Pro	Thuy	Тут 920	Ala	Cln	Ттр	Glu	<i>9</i> er 925	Ma	Glγ	GľΩ
Ile	Lец 930	Thr	Ala	Gły	Leu	Asn 935	Ser	Gìn	gju	Уļя	Asp 940	Ile	Leni	Hi G	Ala
Ptx: 945	Leo	Asp	СЈЛ	Ser	λւ <u>ս</u> 950	िक्य	SIA	Ala	Lem	955 955	Thu-	Тут	፲ ∤፻	Ile	Arg 960
GLn	Val	Ala	Lyb	Pro 965	Ala	Ala	Ala	ll€	Lyn 970	Seor	Arq	Asp	λер	IД: 975	ፒኒጥ
Сſи	Тут	<u>le</u> u	1. 6 0 980	ile	Asp	Aen	Gln	val 985	Ser	Ala	Ala	lle	ъу:: 990	The	Ta
Arg	Ile	Ala 995	Glu	Ala	Ile		\$ж т 1000	IJ¢	G l n	Lou		Val 1005	Asn	Arg	Tha:
Leu 1	G10 (010	ASD	Val	61 0	Gչս :	Asn 1015	Ala	His	Ser		Vа1 1020	IJe	Ser	Arg	Gln
Ptoe 025	lthe	IJe	Asp	Trp	ASP 1030	Lув	Тух	æn		Arg 1035	Тух	Ser	ጥትፈተ		Ala 1040
GΙA	Væ1	Ser	glo Slo	ن د ن 1045	VаЛ	Tyr	Tyr		Glu 1950	yzu	Tyr	Πc		1270 1055	Tha
Met	Arg	Ile :	1060 GJA	Gln	Tht.	lys		Med. 1065	Asp	Ala	ΓEΠ		Glන (070)	Ser	Val
Ser	Gln :	ජනා 1075	Gln	Leu	Asn		Авр 1080	Th ar	Val	GJυ		Ale LDB5	Рю	Юet	Ser
Tyr J	Leu 1090	Thr	Ser.	Phé	Glu	Gl.n 1 09 5	Vel	Ala	Àπŋ		iys Lyb	Val	Ile	Sear	MΑ
Тул 105	Hi:	Æp	Asin	Tie :	Asn 1110	ÀFO	ASP	Gln	Gly S	160 1115	The	ባут	Phé		Gly 1120
Leu	5er	Glu	מבלני.	Asp 1125	Thr	Gly	Glu		тут 1130		Arg	Ser		Аер 1135	Dis
Ser	lye	Ph≘	Ser 2140	ASP	αlу	ГÀЗ		Ala 11 4 5	aźĄ	Aran	Ala		Ser 1350	Glu	Ттр
Ais	lys:	11e 1155	<i>শু</i> নু>	Cyti	Pro	Ile	ASD 1160	Pro	Ίγτ	Arg		Thr 1165	Ile	Ατy	Pro
Va2 :	1 1 3t 1170	ጥ _{ረን} ታ	lys	ජනා	Arg	Leu 1175	Tyr	Leu	Leu		Leu 1180	Glγ	an	Γλε	Clu
11e 185	Tìu	Lyn	Gln	Titu	Gly 1190	Ash	Sex	Cys		G.ly 1195	lyr	GL	Thr		ጥ 1200

- Amp Tyr Arg Tyr Clu iau Lys Lau Ala His Lie Arg Tyr Amp Gly Thr 1205 1215
- Trop Asm Thur Pro Ile Thur Phe Amp Val Amo Glu Lye Ile Ser Lys Leu 1220 1225 1230
- Glu Leu Ala Lys Aso Lys Ala Pro Gly Lou Tyr Cys Ala Gly Tyr Gln 1235 1240 1245
- Gly Glu Asp Thy Lew Lew Val Met Phy Tyr Asm Gln Gln Amp Thr Leu 1250 1255 1260
- Asp Ser Tyr Lys Thr Als Ser Met Gln Gly Leu Tyr Ile She Als Asp 265 1270 1275 1280
- Met Glu Tyr 1ys Asp Met Thr Asp Gly Gln 1yr 1ys Ser Tyr Arg Asp 1285 1290 1295
- Asn Ser Tyr Lye Gln Phe Amp Thr Amn Ser Val Arg Arg Val Asn 1300 1305 1310
- Arg Tyr Ala Glu Asp Tyr Glu Ile Rno Ser Ser Vel Asa Sor Arg Lys 1315 1320 1325
- Gly Tyr Asp Trp Gly Asp Tyr Tyr Leu Ser Met Val Tyr Asn Gly Asp 1330 1335 1340
- The Pro Thr Ile Ser Tyr Lys Als Thr Ser Ser Amp Lou Lys Ile Tyr 345 1350 1355 1360
- The Sor Pro Lys Lea Arg The The His Asm Gly Tyr Glu Gly Gln Gln 1375 \$1375
- Arn Asn Gli Cys Asn Lau Mat Asn Lya Tyr Gly Lya Leu Gly Asp Lya 1380 1385),390
- Pie fle Val Tyr Thr Ser Leu Gly Val Ash Pho Ash Ash Ser Ash 1395 1400 1405
- lys Leu Met Phe Tyr Pro Val Tyr Gln Tyr Asn Gly Amn Val Ser Gly 1410 1415 1420
- Lem Scr Cln Cly Arg Lem Lem Phe His Arg Asp Thr Asp Tyr Ser Ser 425 1430 1435 1440
- Lys Val Glu Ala Trp Ile Pro Gly Ala Gly Arg Ser [co Thr Asm Pro 1445 1450 1455
- Ass. Also Also II.e Gly Asp. Asp. Tyr. Also Thr. Asp. Ser. Let λ 20 Lys. Pro. 1460 1465 1470
- Ash Asp Leu Lye Glu Tyr Val Tyr Met Thr Asp Ser Lye Glu Thr Ala 1475 1485
- Thur Asp Val Ser Gly Pro Val Asp The Asm Thur Ala Jie Sam Pro Ala 1490 1495 1500
- Lys Val Gln Val Thm Val Lys Ala Gly Ser Lys Glu Gln Thr Phe Thr 505 1510 1515 1520
- Ala Asp Lya Asm Val Ser Ile Glo Pro Ser Pro Ser Phe Asp Clu Mot 1525 1530 1535
- ABN Tyr Gln Pha Ash Ala Leu Glu Fle ABP Gly Ser Ser Leu Ash Phe 1540 1545 1550

- The Asn Asn Ser Ala Ser Ile Asp Ile The Phe The Ala Rhe Ala Glu 1555 1560 1565
- Lys Val Ser The Asp Asn Ser Leu Thr Leu Arg Giz Asn Giu Asn Giy 535 1590 1595 1600
- Ala Gin Tyr Met Gin Txp G)y Vel Tyr Arg I): Axg Iku Asn Thr Leo 1605 1616
- Pho Ale Ary Gli Leu Val Ale Arg Ala Thr Thr Gly The Asp Thm The 1620 1625 1630
- Leu Ser Met Glu Thr Gln Asn Ilo Glu Chu Pro Gln Leu Gly Lya Gly 1835 1640 1645
- He Tyr Ala Thr Phe Val Ile Pro Pro Tyr Amn Pro Ser Thr His Gly 1650 1655 1660
- Amp Glu Arg Trop Fine Lym Leu Tyr Ile Lym Him Val Val Amp Amp Agn 665 1670 1675 1680
- Sar Ris Ile Ile Tyr Ser Gly Gln Leu Lys Asp Thr Asn Ile Ser Thr 1685 1690 1695
- Thr Leu Fhe Ile Fro Leu Asp Asp Val Pro Leu Asm Glm Asp Tyr Ser 1700 1705 1710
- Ala Lys Val Tyr Met Thr Phe Lys Lys Ser Pro Ser Asp Cily Thr Trp 1715 1720 1725
- Top Cly Pro His Phe Val Ang Asp Asp Tys Cly Ilo Val Thr Yle Aso 1730 1735 1740
- Pro Lys Ser Ile Leu Thr Ris Phe Glu Ser Val Aen Val Leu Aen Aer 745 1750 1755 1760
- The Ser Ser Glu Pro Net Asp Phe Ser GJy Ala Asa Ser Lau Tyr Phe 1765 1770 1775
- Trp Glu Lau Pho Tyr Tyr Thr Fro Met Leu Val Ala Glo Arg Leu Leu 1780 1785 7790
- His Glu Glin Amn Whe Amp Glu Ala Amn Arg Trp Leu Lym Tyr Vel Trp 1795 1800 1805
- Ser Pro Sor Cly Tyr Ile Val His Gly Ghn Ile Gln Amm Tyr Gln Trp 1810 1815 1820
- Asti Val. Ang Professional Lead Clu Asp Thin Sen Trip Asia Sen Asp Profession 1835 1840
- ABD Ser Vol Asp Pro Asp Ala Val Ala Glu His Asp Pro Met His Tyr 1845 1850 1855
- Tys Val Ser Thr Phe Met Ary Thr Law Asp Tou Law Ilo Ala Arg Gly 1860 1865 1870
- Asp His Ala Tyr Arm Glo Leu Glu Arg Asp Thr Leu Asm Glo Ala law 1875 1880 1885
- Mot Top Tyr Met Gin Ala Leu His Lew Low Gly Asp Dys Pro Tyr Lew

1890	1895	1	900
Pro Leu Ser	מא קדו? שלו: שלו:	Asp Pro Arg Leu .	Asp Lys Alo Ala Asp
905	1.910	1915	1920
No The Thu	Gin Ser Ala Ric	Ser Ser Ser Ile	Val Ala Leu Arg Gln
	1925	1930	1935
	Ala Leu Leu Ser	Leo Arg Sar Ala	Asn Thr Leu Thr Asp
	1940	1945	1950
1au Phe Leu		Glu Val Met Met	Aso Tyr The Gla The
1955		1960	1965
Leu Ala Gln	Arg Val Tyr A sn		Lou Ser Ile Asp Gly
1970	1975		980
Gìn Pro Leu	Tyr Leu Pro Ile	Tyr Ala Thr Pro	Ala Amp Pro Lys Ala
985	1990	1995	2000
Lou Lou Ger	Ala Ala Val Ala	Thu Ser Gln Gly	Gly Gly Lys Leu Pro
	2005	2010	2015
	Mot Ser Lew Trp	Ang Phe Pro His	Met Leo Glo Amn Ala
	2020	2025	2030
Ang Ser Wet		Thr Gin Phe Gly	Ser Thr Leo Glo Aso
2035		2040	2045
lle lle Glu	Arg Gln Asp Ala		Ala Leu Leu Glin Aen
2050	2055		2060
Gln Ala Ala	Glu Lau 11e Lau	Thr Asn Leu Ser	Ile Gìn Amp lym Thr
065	2070	2075	2080
rle Glu Glu	Lau Asp Ala Glu	ilys Ohr Val Leo	Glu bys Ser bys Ala
	2065	2090	2095
	Ser Arg Phe Asap	o Ser Tyr Ser lyc	kao Ris Asp Glo Asn
	2100	2105	2110
11e Asn Ala		i Ala Met Thr Leu	Arg Ala Ser Ala Ala
2115		2120	2125

Asp Leu Val. Pro Asp Ille Phe Clly Phe Ala Clly Clly Gly Ser Arg Trp 145 2150 2150 2155 2160 Clly Ala Ille Ala Glu Ala Thr Gly Tyr Val Net Glu Phe Sur Ala Asp 2165 2170 2175

Gly Leu Thr Thr Ala Val Glm Ala Ser Arg Leu Ala Gly Ala Ala Ala

2135

Val Not Asn Thr Glu Ala Asp Lye Ile Ser Glu Ser Glu Thr Syr Arg 2180 2195 2190

Arg Arg Gln Glu Trp Gln Die Gln Arg Asn Asn Ala Glu Ala Glu 2195 2200 2205

leu bys Gin Leu Asp Ala Gin Leu Lys Ser Leu Ala Val Arg Arg Giu 2210 2215 2220

Als Als Val Leu Gln Lys Thr Ser Leu Lys Thr Gln Gln Glu Gln Thr 225 2230 2235 2240

2130

Gin Ala Gin Lou Ala Pho Lou Gin Arg Lys The Ser Ash Gin Ala Leu 2265 2250 2255

Tyr Ash Trp ion Arg Gly Arg ion Ala Ala Ilo Tyr Has Glo Me Tyr 2260 2270

Asp Leu Ala He Ala Arg Oys Leu Met Ala Glu Gln Ala Tyr Arg Trp 2275 2280 2285

Glu lle Ser Asp Asp Ser Ala Arg Phe Ile Lys Pro Gly Ala Trp Gln 2290 2295 2300

Gly Thr Tyr Ala Gly Leu Leu Ala Gly Glu Thr Leu Mot Leu Ser Leu 305 2310 2315 2320

Ale Gin Met Glu Asp Ale His Lou Arg Asp Lys Arg Ala Leo Glu 2325 2330 2335

Val Glu Arg Thr Val Ser Leu Ala Glu Ile Tyr Ala Gly Leu Pro Gln 2340 2345 2350

Amp Lym Gly Pro Pale Ser Leu Thr Glπ Clu Ilo Glu Lym Leu Val Asn 2355 2360 2365

Ale Gly Ser Gly Ser Ale Gly Ser Gly Asn Asn Leo Ale Pho Gly 2370 2380

Ala Gly Thr Asp Thr Lys Thr Ser Leu Cln Ala Ser Ile Ser Leu Ala 385 2390 2395 2400

Arg Ile Lyn Gin Ile Ser Val Thr Leu Pro Ala Leu Leu Giy Pro Tyr 2420 2430

Gin Asp Vai Gin Ala The Leu Ser Tyr Gly Asp Lys Ala Gly Leu Ala 2435 2440 2445

Asin Gly Cyb Ala Ala Lou Ala Vol Ser His Gly Thr Asin Aqu Sor Gly 2450 2455 2460

Oln Fire Glin Leu Asp the Asn Asp Gly Lys Pho Leu Pro the Glu Gly 465 2470 2475 2480

lle Ala Ile Asp Gln Gly Thr Leu Thr Leu Ser Fhe Pro Asn Ala Ser 2485 2490 2495

The Pro Ala Lye Gly Lys Glm Ala The Met Len Lye The Low Asn Aep 2500 2505 2510

lle lle leu His Ile Ang Tyr Thr Ile Lys 2515 2520

<210> 14

<211> 1481

<212> PRT

<213: Photortabdus luminescers:</p>

<400> 14

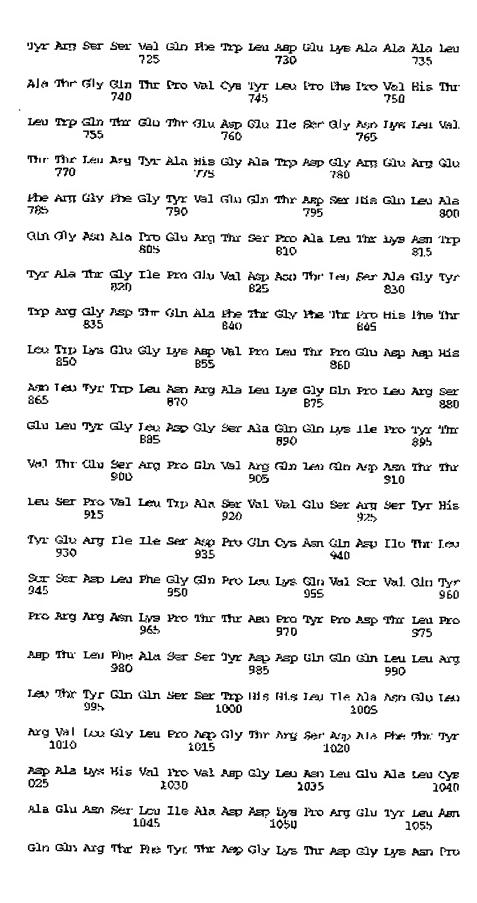
Met Gln Asn Ser Gln Thr Pho Ser Met Thr Glu Leu Ser Leu Pro Lys 1 10 15

Gly Gly Gly Ala 11e Thr Gly Met Gly Glu Ala Leu Thr Pro Ala Gly



			20					25					30		
Pro	Авр	Gly 35	Mert,	Bla	A1a	Leu	Ser 40	Leu	PTO	Leu	Pro	11e 45	ser	Ala	Gly
Arg	Gly 50	Ίγτ	Ala	Pro	Ser	Lец 55	The	Lou	Asn	Тух	Asn 60	Ser	Gly	'191x	Gly
Asn 65	Ser	Fro	Fibe	Gły	Legg 70	Gly	Тър	Д Бр	CAE	Asn 75	Val	Met.	TÎIT	T.le	Л ту 80
Arg	Arg	Thr	Ser	Thr 65	Gly	Val	Pro	Aeto	ग уድ 90	ሊ ነብን	Glu	Thre	Assp	Thr 95	Plac
Leu	Gly	PTO	Glu 100	GЉ	G).u	Val	Leu	Val 105	Val	Ala	Leu	Jeson	Glu 110	Ala	яy
போ	Ala	А вр 115	ıle	Arg	Sea:	G) _{k)}	8ar 120	Seor	Leu	Gln	Gly	11 e 125	Aen	Leu	Gĵy
Met	Tîm: 130	Phe	Tîur	Val	The	Gly 135	тут	Arg	Ser	Arg	I.ен 140	(H) u	S ec	អ្នក	Stoc
Ser 145	Arg	Leo	Glu	Tyr	Trp 150	gln	Pro	Gln.	Thr	That 155	Gly	Aka	'lfre	ASP	She 160
Tro	Lau	Ile	Tyr	Ser 165	Pχυ	Asp	Gly	CJD	Ala 170)ds	Ιæυ	Leu	ദു	Ly:s 175	Asn
Pro	Gln	λ)а	Arg 180	Tle	Ser	aen	Pro	Leu 185	Asn	Val	Aeto	СJD	Thr 190	МZ	Gln
Ттр	Leu	Leu 195	Glu	Me	Sex	Val	Ser 200	Ser	His	Gly	ŒПла	01n 205	Ha	Tyr	Тут
Ġln	Tyr 210	Ary	Ala	Glu	ABP	Glu 215	The	Aap	Cys	Glu	Thr 2 2 0	уер	Glu	Leu	Tìu
Ala 225	Hís	Pro	λεη	Thur	9 ta r 230	Val	Gln) ing	Tyr	Løu 235	Gln	Val	Val	Кiв	1yr 240
αгл	Asm	Leu	Thar	Ala 2 4 5	Sec	Glu	Val	Phe	Pro 250	Thr	Leu	Aso	Gly	Дар 255	Авр
Pro	Less	Lys	Ser 260	ദു	Тф	Leu	Phe	Ω⁄ප 265	Len	Vel	Phe	Ажр	ፕ ኔ፰ 270	$\operatorname{Gl}_{\mathbf{Y}}$	Glu
Arg	Lys	Лет і 275	Ser	Leu	Ser	Glu	Мер 280	Pro	Pro	Fhe	Lув	Ala 285	ימלדי	Ser	Aen
Jrp	ნ ა ს 290	വുള	Arg	Гйа	Asp	л rg 295	Pha	Ser	Arg	Tyr	Glu 300	፲ አኋ	Gly	Phe	Ale
Leu 305	yng	ጥェ	Arg	Arg) թա 310	Cys	Arq	Gln	Tle	1æi 315	Met,	Pho	His	Arg	Leu 320
Gln	าใช	راضا	Ser	Gly 325	מנט	Ala	lys	Gly	Авр 330	аер	Glu	Pica	λla	1 <i>e</i> u 335	Ve I
S1.3T	Arg	Leu	11⊕ 340	Leu	Æp	тук	Asp	Glu 345	Asn	Ala	Val	Val	Ser 350	TÎL	Leu
Val	Ser	val 355	Ахц	Arg	Val	Gly	н <u>ј.</u> s 360	Clu	CJU	Asp	CJA	19u 365	Tîtu"	АŽВ	Val

Ala	Lou 370	Рто	Pro	Leu	ദ്ധ	Lou 375	Ala	ľут	GJv	Pro	Phe 380	Glu	Pro	ផាប	glii
Lys 385	Ala	Leni	Tip	Arg	PTO 390	Xet	лар	Va.1	Leu	Ala 395	Aso	the	Asn	Thr	Ile 400
Glo	Àrg	מנו	ദ്ധ	Leu 405	Lesta	ÀSP	Leu	Gľu	Gly 410	Glu	ദിഴ	Vəl	Prtx	Gly 415	lle
Leu	ΊУΤ	Gln	Авр 420	Lys	Asm	Glγ	Trp	Trp 425	Tyr	Arg	Sec	Ala	Gln 430	ΑIŢŞ	Œľυ
Thr	Gly	Glu 435	Glu	Met	Azo	Ale	Val 440	Thur	Ттр	Gly	Lys	Met 445	Gìn	Fæn	L⊕)
Pro	Ile 450	Thr	Pmo	Ala	Ίle	Gln 455	ASP	Asn	Ala	Sor	Leu 460	Met	Aep	Ile	Aşm
61y 465	Àą	Gly	Gln	Len	љ _ж у 470	Tup	Val	Ile	Thr	ങ്യ 475	Pro	Gly	Leu	Arg	Gly 480
T ያፕ	Hie	Ser	GJŋ	His 485	Pro	Pap	Gly	Ser'	Trp 490	The	A rg	Ehe	Tîre	Pro 49 5	ĭæu
нія	Ala	Lau	₽то 500	Ile	Glυ	Тух	Thir	нів 505	Pro	Arg	Ala	G i n	Leu 510	Ala	qs4
Leu	Met	Gly 515	Ala	Gly	Leu	Ser	Asp 520	Leu	Val	Leu	Ile	Gly 52 5	Pno	Lys	8ecr
Val	Arg 530	Leu	Tyr	Ala	len	Asn 535	Arg	Авр	αз	Ithe	Thr 540	Glu	Gly	Arg	ke p
Val 545	Val	Glŋ	Ser	дλ	Gly 550	Ile	Thu	Leu	PTO	Ն⊖ս 555	Pro	З	ΑĴθ	Asp	Ala 560
Arg	Lyg	Lex	Val	Ala 565	Pho	3er	Авр	Nal	1 <i>e</i> 0 570	Oly	Scr	Gly	Gln	Ala 575	His
Leu	Val	Glu	Va1 580	Ser	Ala	Thr	Lys	Val 505	Than	Cys	тър	Pη	Asa 590	Lou	GLY
Kis	Gly	Arg 595	Phe	GLV	Gln	Pro	T)@ 600	The	Leo	Pro	αlγ	9he 605	Ser	Gln	Ser
Ala	ለጋል 610	Aan	Fhe	Asar	Pro	А э р 615	Arg	Val,	His	Lou	Ala 620		Leu	ABŢ	Gly
Ser 625	Gly	Pro	Ala	Asp	Le n 630	Tle	Tyr	٧al	His	Мв 635	እሟን	Юя	Lex	Aæp	Tle 640
Phe	Ser	Aan	Glu	Ser 645	G1y	Aan	Сју	Phe	Ala 650	Сľъ	Pro	Phe	Tra	Leu 655	Arg
Phe	Рго	P SP	660 660	kou	Àrg	Phe	Авр	ሎ _ዚ) 665	Ttn-	ζув	Gln	נשנ	G)n 670	Val	Ala
увр	Va l	Cln 675	Gly	Legy	Cly	Val	Val 680	Sea:	T ajs ju	IJ.c	Lew	Ser 685	Val	Pro	Нis
Met.	Ala 690	Pro	His	Hie	Ттр	Arg 695	Cys	Аер	Leu	Thr	Ле п 700	Ala	Lys	PTO	TIP
1.eu 705	ī.eu	Ser	ദിവ	Net	æn 710	Asn	Asm	Met.	Gly	л]д 715	His	Hìs	Thur	Leu	His 720



		;	tiaor				:	L065				;	1070		
Mur	Pro	Len 1075	Lγς	Tra-	Pro	Mar	Arg 1080	ദ്ധ	Ala	Leu		Ala 1085	₽'næ	Thr	Glu
ייינר דער	Ala (090	Val	Leu	Tir	ങ്ങ 1	Sur 1095	Lena	leu	Seur		Ina 1100	Аэр	Gly	Gly	Ile
The 105	Pno	Asp	Glα	Leu	Pro	Gly	Leni	Leni		Gln 1115	Ala	Gly	Тут		Gln 1120
Glu	כתל	Тут	Leu	Phe 1025	Pro	Len	&:τ°	Gly 2	Glu 1130	क्षेत्रा	Gln	Val		Val L135	Ala
Arg	īys	gyy	ТУл: \$140	Thr	Ģap	Тут	Gly	Thu- L145	Glu	Val	GÌr		Մ որ L150	Arg	Pro
Val	Ala	Gl n 1155	Arg	Asn	Trac	Gln :	Leu (260	The	Gly	Lys		Thr 1365	Leu	Lye	тър
Asp 1	Ծիր L170	His	Тут	ĽVB		Ile 175	Thi:	Gin	ፓ ትነሶ		љу, 180	AlA	λla	Gjy	Leu
Thr 125	АЭJ	Ser	Ala	nen !	ብን። 1350	Аэр	Тгр	Arg	Phe	ւթս 1195	'Ilm	Pro	Met	GJD (Leu L2011
Thar	yesto	Ile	A so	Asp 1205	Asn	V2)	Hie		1 le 1210	Thr	Leni	жр		1.e.) 1.215	@] Y
ÀFQ	Pro	Val	Tir 12 2 0	ලාා	Arg	Phe	TYP	Gly L225	Iļe	Głu	Aan		val L230	Ala	Thr
Gly	Tyr	Ser 1235	විතා	jao	Glu		Lye: 1240	Pro	Phe	Shir		P170 1265	Va1	yesto	Val
V:n	A1a 1250	Ala	lle	Ala	Leu 1	Tha: .255	СЛУ	Pro	Len		Val (260)	Ala	сји	СÀв	Leu
Val 265	Tyr	Ala	Pro	Asp I	Ser L270	Ίτρ	M≘t	Pro	Leu	Phe 1275	ĠĴУ	Gin	Gl.u		Phe L280
Asn	Tim	Leu	Thr	G) 63 1285	Glu	Glu	Gln		Thr 1290	Leu	Arg	ДSP	Leu 1	λ rg 1295	Ile
He	The	Glu	Аер 1300	Ттр	Arg	Ile	-	АЗа 1305	Leu	Ala	Arg		Arg 1310	2,tT>	Leu
බො	Ser :	Gin 1315	Lys	Ala	Gly	Thr	Pno 1320	T ಜ ಬ	Vel	Lys		Leu 1325	Thr	Asn	Sor
ıle	Gly 1330	Len	Pro	Pro	l His	Ae n .335	Leu	Xet.	I,ee)		Thr 1.340	Asp	Ary	Tyr	ASO
Arg 345	As _D	Ser	Glu	д Сјп	ផា 1350	Ile	Àгŋ	Gln		Val 1355	Ala	#he	5er		Gly 13 6 0
Phe	Gly	Arg). J	Leu 1365	ផាក	Ala	Ala		A119 (370)	Riø	Glu	Ala		Glu 1375	Ala
לגו,	Gln	arg	A :an 1360	Gìn	Asp	Gly		Ն։ 13B5	Vel	TÎTE	lys		GJ.u 1390	Aup	Thr
Ĺ₩≅	lhr	Arg 1395	Τηγ	Ala	Ilc		Gly 1400	Arg	Thir	Glu		л ар 1405	Aen	Lys	Gly

Glo Ala 1le Arg Thr Tyr Gln Pro Tyr Phe Leu Arn Arn Trp Arg Tyr 1410 1415 1420	
Val Ser Asp Asp Ser Ala Arg Lye Glu Ala Tyr Ala Asp Thr His The 425 1430 1435 1440	
Tyr Asp Pro Ile Gly Arg Glo Ile Cln Val Ile Thr Ala Lys Gly Trp 2445 1450 1455	
Leu Arq Gln Agn Gln Tyr Phe Pro Trp Phe Thr Val Ser Glu Agp Glu 1460 1465 1470	
Asn Asp Leu Ser Ala Asp Ala Leu Val 1475 1480	
<210> 18 <211> 23 <210> Depa <213- Artificial Sequenco	
<220: <223> Description of Artificial Sequence:oligonucleotide	
<600> 15 cgggatccga tgettittaaa agg	3
<210> 16 <211> 16 <212> DW <213> Artificial Sequence	
<2205 <223> Description of Arthficial Seguence:oligonucleotide	
<pre><400> 16 gogreatigs fittens</pre>	6
<pre><210> 17 <211> 19 <212> INA <213> Artificial Sequence:</pre>	
<220> <223> Description of Artificial Sequence:oligonuclootide	
<pre><400> 17 cattagraget ogaangten</pre>	9
<210> 18 <213> 26 <212> DWA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonocloptiche	
<400> 18 gegegagete ttachtsatg gtgCag 2	6
<210> 19	

<211> <212> <213>		
<220> <223>	Description of Artificial Sequenca:oligonucleotide	
<400>	19 Aggic caigcagaai irgcanac	28
		•
<210><211><211><212>	18 IXA	
~213 >	Artificial Sequence	
<220> <223>	Description of Artificial Sequence:oliganucleotide	
<400>	- '	
ggcaat	Ogica godatasg	10
<210> <211> <212> <213>	18	
<220> <223>	Description of Artificial Sequence:oligonuclectide	
<400> cette	ži segos ggsagage	18
<210> <211> <212> <213>	26	
<220> <223>	Description of Artificial Segmence:oligonucleotide	
<400> gaceto	22 gagt tamacgagng ogtoag	26

		,

PCT





INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C12N 15/31, 15/82, 15/10, 1/21, 5/10, A01H 5/00, C07K 14/24, A01N 63/02 **A3**

(11) International Publication Number:

WO 99/42589

(43) International Publication Date:

26 August 1999 (26.08.99)

(21) International Application Number:

PCT/EP99/01015

(22) International Filing Date:

18 February 1999 (18.02.99)

(30) Priority Data:

09/027,080 60/116,439 20 February 1998 (20.02.98) US

20 January 1999 (20.01.99) US

(71) Applicant (for all designated States except AT US): NOVAR-TIS AG [CH/CH]; Schwarzwaldallee 215, D-4058 Basel (CH).

(71) Applicant (for AT only): NOVARTIS-ERFINDUNGEN VER-WALTUNGSGESELLSCHAFT MBH [AT/AT]; Brunner Strasse 59, A-1235 Vienna (AT).

(72) Inventors; and

(75) Inventors/Applicants (for US only): KRAMER, Vance, Cary [US/US]; 608 Dana Court, Hillsborough, NC 27278 (US). MORGAN, Michael, Kent [US/US]; 5805 Garrett Road, Durham, NC 27707 (US). ANDERSON, Ame, Robert [US/US]; 1005 Green-Pace Road, Zebulon, NC 27597 (US). HART, Hope, Prim [US/US]; 4106 Planters Glen Court, Fuquay-Varina, NC 26526 (US). Warren, Gregory, Wayne [US/US]; 324 Bond Lake Drive, Cary, NC 27513 (US). DUNN, Martha, M. [US/US]; 6201 Oakview Court,

Hillsborough, NC 27278 (US). CHEN, Jeng, Shong [-/US]; 302 Orchard Lane, Chapel Hill, NC 27514 (US).

(74) Agent: BECKER, Konrad; Novartis AG, Corporate Intellectual Property, Patent & Trademark Dept., CH-4002 Basel (CH).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, Cl, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:

23 December 1999 (23.12.99)

(54) Title: INSECTICIDAL TOXINS FROM PHOTORHABDUS

(57) Abstract

Novel nucleic acid sequences isolated from Photorhabdus luminescens, whose expression results in novel insecticidal toxins, are disclosed herein. The invention also discloses compositions and formulations containing the insecticidal toxins that are capable of controlling insect pests. The invention is further drawn to methods of making the toxins and to methods of using the nucleotide sequences, for example in microorganisms to control insect pests or in transgenic plants to confer insect resistance.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

Al.	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
ΛZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
ВВ	Barbados	GH	Ghana	MG	Madagascar	ТJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
ВJ	Benin	ΙE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of Americ
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JР	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		. •
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

Inter- itional Application No 01015 Pui/EP

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/31 C12N15/82

A01H5/00

CO7K14/24

C12N15/10 A01N63/02

C12N1/21

C12N5/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N A01H C07K A01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUM	NTS CONSIDERED TO BE RELEVANT	
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 17432 A (WISCONSIN ALUMNI RES FOUND) 15 May 1997 (1997-05-15) the whole document, particularly SEQ ID NOS 31,46,47,48,49,50,51,60	1-3,7-9, 11-24, 26-36
P,X	WO 98 08932 A (DOW AGROSCIENCES LLC; WISCONSIN ALUMNI RES FOUND (US)) 5 March 1998 (1998-03-05) see pages 209-210,215-224,231-237, and 243-245. /	1-3,7-9, 11-24, 26-36

X Further documents are listed in the continuation of box C.	X Patent family members are listed in annex.
Special categories of cited documents: 'A" document defining the general state of the art which is not considered to be of particular relevance 'E" earlier document but published on or after the international filing date 'L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) 'O" document referring to an oral disclosure, use, exhibition or other means 'P" document published prior to the international filing date but later than the priority date claimed	To later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family
Date of the actual completion of the international search	Date of mailing of the international search report
20 October 1999	0 8. 11. 99
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentiaan 2	Authorized officer
NL - 2280 HV Rijawijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Maddox, A

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int	tional Application No	
C.	T/EP 99/01015	

ssages Relevant to claim No.	uation) DOCUMENTS CONSIDERED TO BE RELEVANT Citation of document, with indication, where appropriate, of the relevant passag	C.(Continua Category *
dus	DAVID JOSEPH BOWEN: "Characterization a High Molecular Weight Insecticidal Protein Complex Produced by the Entomopathogenic Bacterium Photorhabdus luminescens (Nematodes, Biological Control)" THESIS UNIVERSITY WISCONSIN, 1 May 1995 (1995-05-01), XP002076022 see chapter 3	A
ORG 1-36 T RAY)	WO 95 00647 A (COMMW SCIENT IND RES ORG ;SMIGIELSKI ADAM JOSEPH (AU); AKHURST I 5 January 1995 (1995-01-05) the whole document	Α
om a	SZITTNER, R., ET AL.: "Nucleotide sequence, expression, and properties or luciferase coded by the lux genes from terrestrial bacterium" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 27, 1990, pages 16581-169 XP002119674 figure 5	Α
12-19, 29-34	WO 93 07278 A (CIBA GEIGY AG) 15 April 1993 (1993-04-15) the whole document	A
	WO 98 08388 A (MORGAN JAMES ALUN WYNNE; JARRETT PAUL (GB); ELLIS DEBORAH JUNE (GB) 5 March 1998 (1998-03-05) see SEQ ID NO:1	P,A
	; JARRETT PAUL (GB); ELLIS DEBORAH JUNE (GB) 5 March 1998 (1998-03-05)	P,A



PCT/EP 99/ 01015

Box I Obs rvations wher certain claims were found unsearchable (Continuation fitem 1 of first shiet)
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2. Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. X Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
B x II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows:
see additional sheet
As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark on Protest The additional search fees were accompanied by the applicant's protest. X No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 4,5,6,10,25 all completely, and 1-3,12-24, 27-36 all partially

Nucleic acid molecule comprising the claimed regions of sequence ID 1, chimeric genes and hosts containing said molecule, toxins expressed by said regions, and method for producing said toxins and controlling insects using said toxins, method for mutagenizing said nucleic acid molecules.

2. Claims: 7-9,11,26 all completely, and 1-3,12-24, 27-36 all partially

Nucleic acid molecule comprising the claimed regions of sequence ID 11, chimeric genes and hosts containing said molecule, toxins expressed by said regions, and method for producing said toxins and controlling insects using said toxins, method for mutagenizing said nucleic acid molecules.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210
Continuation of Box 3.
The reference to claim 44 in claim 30 is inconsistent with the numbering of the claims, since claim 44 has not been filed. For the purpose of defining the search, claim 30 has been considered to refer to the toxin of claim 20, and searched accordingly.
\cdot

INTERNATIONAL SEARCH REPORT

intr tional Application No

	ent document in search report		Publication date		atent family nember(s)	Publication date
HO	9717432	A	15-05-1997	AU	1050997 A	29-05-1997
WU	9/1/432	A	13-03-1337	BR	9606889 A	28-10-1997
				CA	2209659 A	15-05-1997
				EP	0797659 A	01-10-1997
				HÜ	9900768 A	28-06-1999
				PL	321212 A	24-11-1997
				SK	93197 A	06-05-1998
				AU	2829997 A	19-03-1998
				WO	9808932 A	05-03-1998
				WU 		
WO	9808932	Α	05-03-1998	ΑU	1050997 A	29-05-1997
				ΑU	2829997 A	19-03-1998
				BR	9606889 A	28-10-1997
				CA	2209659 A	15-05-1997
				EP	0797659 A	01-10-1997
				HU	9900768 A	28-06-1999
				PL	321212 A	24-11-1997
				SK	93197 A	06-05-1998
				WO	9717432 A	15-05-1997
wn	9500647	Α	05-01-1995	AU	675335 B	30-01-1997
NO	J300047	,,	00 01 1550	AU	6991694 A	17-01-1995
				EP	0705340 A	10-04-1996
				JP	9500264 T	14-01-1997
WO	9307278	- -	15-04-1993	US	5625136 A	29-04-1997
NO	3307270	^	15 01 1555	AU	2795292 A	03-05-1993
				BG	98747 A	28-02-1995
				BR	9206578 A	11-04-1995
				CA	2120514 A	15-04-1993
				CZ	9400769 A	15-03-1995
				ÉP	0618976 A	12-10-1994
				HU	68261 A	28-06-1995
				JP	7500012 T	05-01-1995
				RO	110263 A	30-11-1995
				SK	37894 A	05-10-1994
				US	5859336 A	12-01-1997
		Α	05-03-1998	AU	4024997 A	19-03-1998
WO	9808388					